Brain transcriptome analyses of paternal-specific mutant GRB10KO and potential insights into dominance behaviour

Al Zadjali, Abduljalil

Award date:
2018

Awarding institution:
University of Bath

Brain transcriptome analyses of paternal-specific mutant GRB10KO and potential insights into dominance behaviour

Abdul Jalil Al Zadjali

A thesis report submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

March 2018

Declaration of any previous submission of the work

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

Candidate's signature

Declaration of authorship

I am the author of this thesis, and the work described therein (including the formulation of ideas, design of methodology, experimental work, and writing) was carried out by myself personally, with guidance from my supervisory team with the exception of the following contributions:

Chapter 2

Dr Kim Moorwood and staff from the animal house helped in setting the crosses and looked after the mice used in the study until they reached the ages required for my study. Dr Kim Moorwood also guided me during mouse genotyping (10% contribution to chapter's data). Dr Humberto Gutierrez helped with the dissection to obtain the samples of the cortex and the subcortical areas from one mouse used in the study (2% contribution to chapter's data).

Chapter 5

Ms Mayra Ruiz provided training and help in carrying out the co-expression network analysis (20% contribution to chapter's analyses).

Chapter 6

Ms Saminathan and Dr Kim Moorwood set the crosses, carried out the collection of liver-tissue, genotyping, RNA extraction and purification. Ms Saminathan also sent samples for sequencing (30% contribution to chapter's data).

Candidate's signature

Table of Contents

Acknowledgements....vi
Summary....vii
List of Abbreviations....viii
Chapter I. Introduction....1
Growth factor receptor-bound protein 10 (Grb10)....2
Grb10 and disease....4
Grb10 and Central Nervous System....4
Structure of Grb10 and the GRB10 pathway....7
Aim....14
Chapter II. Experimental protocols and sample collection....15
Introduction....16
Methods....16
Ethics Affirmation....16
Mice for the experiment....16
Stain base for β-galactosidase (*LacZ*) assay....17
Developmental stages sampling....18
Brain dissection....18
Sample storage, RNA extraction and sequencing....21
Genotyping....21
Results....23
Allele-specific Grb10 gene expression in the brain and heart....23
Sample collection....25
Discussion....26
Chapter III. Transcriptome profile analyses of the Grb10 paternal $KO^{+/p}$....29
Introduction....30
Differential gene expression....32
Methods....32
Grb10 and β-geo Primers....32
Quality Analysis using FastQC....33
Transcript annotation....33
Differential expression analysis....34
Multidimensional scaling plots (MDS) and Heatmap....34
Functional enrichment analysis....34
Results....35

Inserted cassette is only detected in Grb10KO$^{+/p}$ tissue samples 35
High quality RNA-Seq data allowed transcriptome profiling of Grb10KO$^{+/p}$ mice 39
Overall similarity in Grb10KO and wild-type transcriptome profiles 43
Gene ontology enrichment analysis 54
Discussion 64
Chapter IV. Alternative splicing alterations in the Grb10KO$^{+/p}$ 67
Introduction 68
Alternative splicing in the brain 68
Grb10 alternative splicing transcripts 69
Transcriptome analysis of alternative splicing events 70
Tools for computational analyses of transcriptional alternative splicing 71
Aim 73
Methods 73
Analyses of alternative splicing transcripts 73
Results 74
Discussion 78
Chapter V. Changes in the regulatory architecture of Grb10KO$^{+/p}$ mice 82
Introduction 83
Gene Co-expression network analysis 83
Differential Co-expression network analysis 85
Co-expression in the brain 86
Co-expression of Grb10 87
Aim 88
Methods 88
Statistical analysis of co-expression 88
Co-expression structure clustering analysis 88
Differential Co-expression network analysis 89
Bar plots and Heatmap 89
Gene Ontology enrichment analysis 90
Results 91
Distinct cellular functions are enriched in differentially co-expressed clusters 97
Discussion 112
Chapter VI Comparing maternal and paternal Grb10KO 116
Introduction 117
Aim 118
Methods 118

Sample collection and genotyping 118
RNA extraction 119
RNA Quality Control 119
Library preparation and RNA Sequencing 119
Transcriptome profile analyses of the Grb10 maternal $KO^{m/+}$ 119
Results 120
Transcriptome annotation of Grb10 maternal $KO^{m/+}$ and wild-type controls 120
Maternal Grb10$KO^{m/+}$ leads to differential expression of over 100 genes 121
Maternal and paternal KO show distinct transcriptional signatures 122
Maternal and paternal Grb10 associated with distinct functional categories 125
Discussion 131
Chapter VII Final Discussion 133
Introduction 134
Future work: 138
In conclusion: 140
References 142
Supplementary material 172

Acknowledgements

My sincere gratitude to my family, especially my mother, for their moral support.

I also would like to thank my supervisors Dr Araxi Urrutia, Prof Andrew Ward, Dr Kim Moorwood and Dr Humberto Gutierrez who have supported me with their discussions, guidance, opinions and advice were given to me during my studies.

I am grateful to Dr Atahualpa Castillo, Kathryn Maher, Paola, Ricardo Cerda, and Mayra Ruiz and all other past and present 1.29 lab members for their help and advice.

Finally, the work presented here would not have been possible without the economic support of his Majesty *Qaboos bin Said Al Said* Sultan of the Sultanate of Oman. I acknowledge the funding that allows me to take part in this irreplaceable experience of studying a degree in the United Kingdom.

Summary

Variation of gene expression in the brain may play a critical role in determining behavioural phenotypes. Grb10 gene is an imprinted gene expressed from embryonic stages into adulthood. In mice, Grb10 parental alleles are expressed in a tissue-specific manner, with the expression of the maternal allele in the placental trophoblasts and most embryonic tissues. The paternal allele is expressed in certain regions of the brain. Grb10 expression in the adult is limited to few tissues with the maternal allele expressed in muscle tissues and the paternal copy maintains its expression in the midbrain. While maternal knockdown of Grb10 is associated with foetal growth, a function shared with several other imprinted genes, knocking down of the paternal allele mice is associated with increased social dominance. The interactors of Grb10 in the brain, however, are not yet known.

In this project, I aimed to uncover the potential molecular interactors mediating Grb10's function in the brain.

For this, I obtained whole genome transcriptional profiles using Illumina RNA-Seq technology for the midbrain where paternally inherited Grb10 is expressed in adult stages and the neocortex where Grb10 is not expressed for mutantGrb10KO$^{+/p}$ and wild-type. Using differential gene expression analyses, I identified a set of genes altered in the paternal mutant Grb10KO$^{+/p}$. I further examined how these sets of genes change over time by examining samples taken at different developmental stages. Gene expression profile changes in the Grb10KO in the brain are distinct from gene expression changes in the liver suggesting that the maternal and paternal alleles are associated with different sets of genes in different tissues. Co-expression analyses revealed significant shifts in gene to gene relationships in the KO context. Together, my results provide insights into the molecular interactors of Grb10 in the brain and may give clues as to how Grb10 influences social dominance.

List of Abbreviations

+/+	Wild-type
1M Tris-HCl	1Molar Tris-hydrochloric acid
1M	1 Month old
1W	1 Week old
3M	3 Months old
6M	6 Months old
aa	Amino acid
Ad-Grb10	Activated domain of Grb10
Akt	protein kinase B
BGI	Beijing Genomics Institute
BMP	Bone morphogenetic protein
bp	Base pair
BPA	Bisphenol A
c-kit	Receptor tyrosine kinase
CNS	Central nervous system
DICER	Differential Correlation in Expression for meta-module Recovery
DiffCoEx	Differential Co-expression
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphates
E18.5	Embryo 18.5 days
edgeR	Empirical Analysis of Digital Gene Expression Data in R
EDTA	Ethylene Diamine Triacetic Acid
F primer	Forward primer
FDR	False discovery rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GB	Gigabytes
GO	Gene Ontology
GOID	Gene ontology identifier
GTPase	Guanosine triphosphate hydrolase enzymes
HEK 293 cells	Human Embryonic Kidney 293 cells
hGrb10	Human Grb10
ID	Identifier
IGF-1	insulin-like growth factor 1

IGF-2	insulin-like growth factor 2
IGF1R	insulin-like growth factor 1 receptor
Kv1.3	Shaker voltage-gated potassium channel
Kv	Voltage activated potassium channels
LacZ	Neomycin2 gene-trap (a fusion gene formed from the β-galactosidase gene and the neomycin- resistance gene)
LexA-IGFIR beta	LexA DNA-binding domain fused to the intracellular portion of the IGF-I receptor
LEU2	Leucine
MAPK	Mitogen-activated protein kinase
MDS	Multidimensional scaling
MEK1	Mitogen-activated protein kinase 1
mGrb10	mouse Grb10
MHC	Major histocompatibility complex
Mig-10	Abnormal cell migration protein 10
ml	Millilitre
mM	Millimolar
MQ	Milli-Q
n	Number
NaOH	Sodium hydroxide
Nedd4	Neural precursor cell expressed developmentally down-regulated protein
Nedd4-2	Neural precursor cell expressed developmentally down-regulated protein-2
nShc	Neuronal Src homology and collagen
P	Proline rich motif
pA	Polyadenylation
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
Ph	Concentration of hydrogen ion
QC	Quality control
qPCR	Quantitative polymerase chain reaction
Raf1	Rubisco accumulation factor 1
Rho	Ras-homologous
RM1	Rodent Maintenance diet
RNA	Ribonucleic acid
RNA-Seq	RNA-Sequencing

R primer	Reverse primer
SA	Splice acceptor
SE	Standard error
SH2	Src homology 2
SMAD	Homologs of both the Drosophila protein, mothers against decapentaplegic (MAD) and the Caenorhabditis elegans protein SMA (from gene sma for small body size)
Taq	Thermus aquaticus
TMM	Trimmed Mean of M-values
TPM	Transcripts per million
UCSC	University of California Santa Cruz
VEGF-R2	Vascular endothelial growth factor receptor-2
WGCNA	Weighed gene co-expression network analysis
WT	Wild-Type
Wnt	Wingless-type MMTV (mouse mammary tumour virus) integration site
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
β-geo LacZ	neomycin2 gene-trap (a fusion gene formed from the β-galactosidase gene and the neomycin- resistance gene)
µg	Microgram
µl	Microliter
µM	Micromolar

Chapter I. Introduction

Growth factor receptor-bound protein 10 (Grb10)

Grb10 gene encodes an intracellular adaptor protein known as growth factor receptor-bound protein. Grb10 has been extensively studied in human and mouse models as it is a major regulator of growth during development. Grb10 expression has been detected in almost every tissue in the mouse embryo including the brain (Charalambous et al., 2003; Hikichi et al., 2003; Garfield et al., 2011). Interestingly, Grb10 is an imprinted gene (see Box 1) and in both human and mouse show different expression profiles for each of the alleles in a single individual depending on the parent of origin and the differential expression of allele copies can be mediated through methylation of DNA and modification of histone (Feil & Berger, 2007).

In mice, maternal allele-specific expression of Grb10 occurs in placental trophoblasts and in most foetal tissues (Charalambous et al., 2003, Garfield et al., 2011). This maternally expressed Grb10 ($Grb10^{m/+}$) is dramatically down-regulated in late gestation and postnatally it is expressed in only a subset of peripheral tissues (Smith et al., 2007). In contrast, in humans, biallelic expression occurs in most foetal tissues, but as in mice, the maternal allele-specific expression occurs in placental trophoblasts (Charalambous et al., 2003; Smith et al., 2007; Monk et al., 2009; Charalambous et al., 2010). Maternal Grb10 in the mouse suppresses the growth of the placenta and the offspring (Charalambous et al., 2003; Smith et al., 2007; Charalambous et al., 2010) hence protecting the pregnant mouse to maintain its resource and any other complications of the pregnancy. It acts as a mediator of nutrient supply that expressed in the mother and as a demand which expressed in the offspring regulating the proportions of lean mass and fat tissue during offspring development hence influencing energy homeostasis in adulthood (Charalambous et al., 2003; Cowley et al., 2014; Smith et al., 2006; Charalambous et al., 2010).

According to Nantel et al. (1999) most of the Grb10 molecules peripherally associated with mitochondria, few of them can be found at the plasma membrane and in actin-rich membrane ruffles depending on the growth conditions, overexpression of Grb10 leads to its mislocalization to the cytoplasmic matrix. It may be involved in communication between outer membrane of the mitochondria and the plasma membrane.

Box 1. Imprinted genes

Imprinted genes are those where the paternal and the maternal alleles are expressed in a different manner (figure B.1). Imprinting evolves in conditions where there is high degree of sexual conflict where optimal female investment in embryos and pups differs according to the type of the sex in a species. Imprinted genes have been described in insects (Pseudococcusnipae) (Schrader Franz, 1921), plants (in maize) (Kermicle, 1970) and mammals (in mice) (Lyon & Glenister, 1977). To date, nearly 150 imprinted genes have been reported in humans and mice (Barbaux et al., 2012). Most of these are involved in regulating embryonic and placental growth, and postnatally, in the maintenance of metabolic homeostasis (Ishida & Moore, 2013) and the behaviour (Garfield et al., 2011).

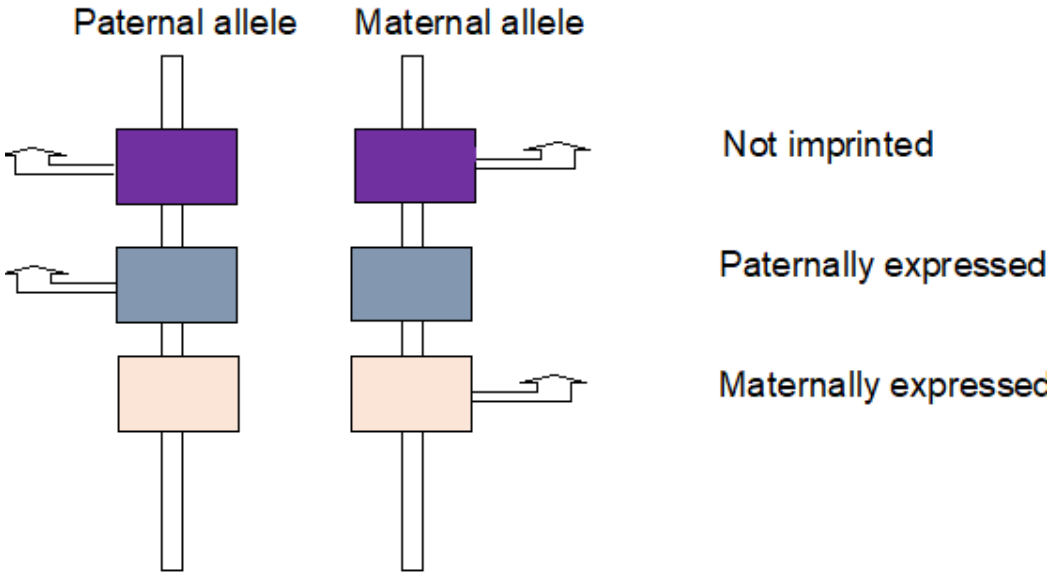


Figure B.1. Mode of expression in biallelic and monoallelic (imprinted) genes. In biallelically expressed genes (purple colour) paternal and maternal inherited alleles are expressed. Imprinted genes, exhibit monoallelic expression where only the maternal (pink) or the paternal (blue) allele expressed. Arrows indicate transcriptional activity.

Grb10 plays a role in cell proliferation and regulation of apoptosis. In spite of its roles in pathways activated by insulin receptor (IR) and insulin-like growth factor receptor (IGF-R) contradictory results shown in some studies whether Grb10 has inhibitory or stimulatory effects. As a study done by Morrione et al., (1997) indicated that stable overexpression of Grb10 inhibits IGF-1 mediated cell proliferation. The inhibitory

effect on IR and IGF-R signalling has also been reported by other investigators (Liu & Roth, 1995; Stein et al., 2001; Wick et al., 2003). On the other hand, a study was done by O'Neill et al., (1996) reported that the expression of Grb10 has a stimulatory effect on insulin and IGF-1. This is also agreed by Wang et al., (1999) who showed that overexpression of Grb10 increased DNA synthesis upon the growth factor such as insulin, or insulin-like growth factor 1 (IGF-1) and platelet-derived growth factor-BB. The contradictions might be due to experimental procedures or the use of different cell lines or due to using different Grb10 isoforms, that perhaps have different functions and may compete for common substrates (Morrione, 2003). In the regulation of apoptosis Nantel et al. (1998) reported that expression of mutant Grb10 induced apoptosis of the cell while the co-expression of wild-type Grb10 increases cell survival.

Grb10 and disease

The genetic expression can be vulnerable to a specific disease condition, and it is the most fundamental level in which a genotype gives rise to a phenotype, as these expressions can be influenced by the environment effect of the interaction between the genes. Research studies showed that Grb10 is associated with Silver-Russell syndrome (SRS) disease in human which is a growth disorder characterized by prenatal and postnatal growth failure, and other abnormal features, due to maternally overexpression of the imprinted Grb10 gene, rather than the paternally expression (Kotzot et al., 1995; Joyce et al., 1999; Monk et al., 2000). Yoshihashi et al. (2000) also suggested that the maternally transmitted mutant Grb10 impacted to the SRS. These results were argued by Hitchins et al. (2001) and McCann et al. (2001) as they considered Grb10 is unlikely contributes to SRS.

Grb10 and Central Nervous System

Unlike most imprinted genes where only the maternal or the paternal allele is expressed, Grb10 is considered unique among imprinted genes for its reciprocal imprinting pattern and it is the first example of an imprinted gene that regulates social behaviour that exhibits monoallelic expression in a tissue-specific manner. Grb10’s paternal allele is expressed in a subset of neurons in the brain during embryonic

development in the mouse (Yamasaki-Ishizaki et al., 2007; Sanz et al., 2008; Garfield et al., 2011; Plasschaert & Bartolomei, 2015) in most subcortical regions of the brain and the spinal cord. Hikichi et al. (2003) reported that a brain-specific promoter mainly drives transcription of Grb10 in the brain with minimal transcription from the major promoter and the transcription of Meg1/Grb10 in the mouse brain starts from 9.5 days post-coitum (d.p.c). Later studies (Sanz et al., 2008; Garfield et al., 2011) showed that expression in the brain occurs at ages between E11.5 and E14.5 (Figure 1.1). There is a very limited expression of the maternally inherited copy expressed in the epithelium of the choroid plexus, ventricular ependymal layers and the meninges during late embryonic development (Garfield et al., 2011). Interestingly, disabling paternal Grb10 not only involved in postnatal social behaviour but also brain size development (Garfield, 2007) suggesting that Grb10 expression in the brain regulates variable functions.

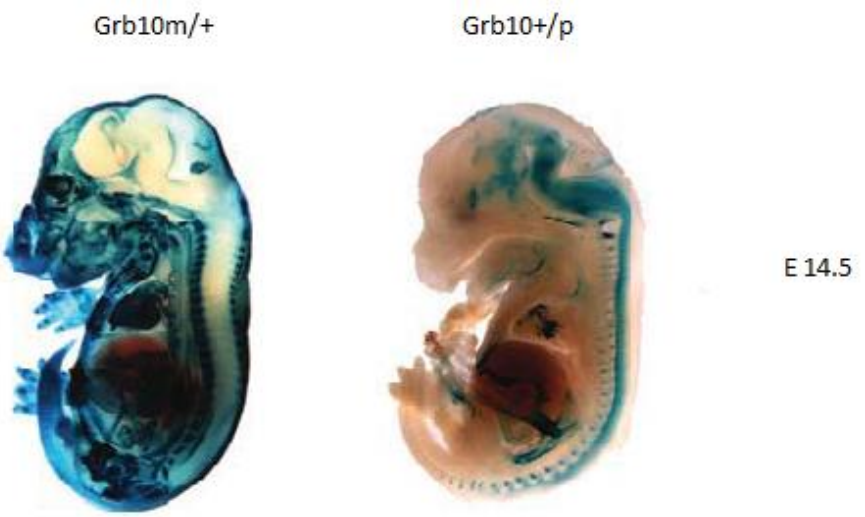


Figure 1.1. Embryonic gene expression pattern of Grb10 *LacZ* reporter for Grb10 expression of maternally (left) and paternally (right) inherited alleles at E14.5 in hybrid strain mice (C57BL/6 with CBA). The maternal allele expressed in tissues of mesodermal and endodermal origin, with expression in the brain limited to the epithelium of the choroid plexus, ependymal layers and the meninges. The paternal allele expressed in the diencephalon, ventral midbrain and the medulla oblongata extending caudally along the ventral spinal cord with few discrete sites showing low-level expression seen in other tissues (Adapted from Garfield et al., 2011).

Interestingly, whereas most imprinted genes are mostly expressed during embryonic and juvenile stages, the expression of the Grb10 paternal allele in brain areas continues into adulthood in both human and mice (Hikichi et al., 2003; Monk et al., 2009). Garfield et al. (2011) revealed that knocking out expression of paternal Grb10

increased social dominance in mice which more likely to carry out allogrooming (barbering or removing whiskers from other mice) behaviour with the cage mates, and this behaviour has been previously considered to be associated with social dominance (Sarna et al., 2000). These mice were also more likely to exhibit a higher degree of social dominance in the tube test (Garfield et al., 2011). This test was first described in 1961 as a measure of social dominance (Lindzey et al., 1961) and is now widely used as a measure of this trait (Wang et al., 2011; Filiano et al., 2013; de Esch et al., 2015; Arrant et al., 2016). It consists of placing mice at either end of a clear tube and then observed which mice back off from the tube during these forced encounters. Previously socially isolated unfamiliar adult Grb10KO$^{m/+}$ and/or Grb10KO$^{+/p}$ with WT mice matched for body size were significantly less likely to back down than WT this was not the case for Grb10KO$^{m/+}$ (Figure 1.2) (Garfield et al., 2011).

Because of restrictions in the license for the project, I was unable to carry out behavioural tests of dominance such as the tube test.

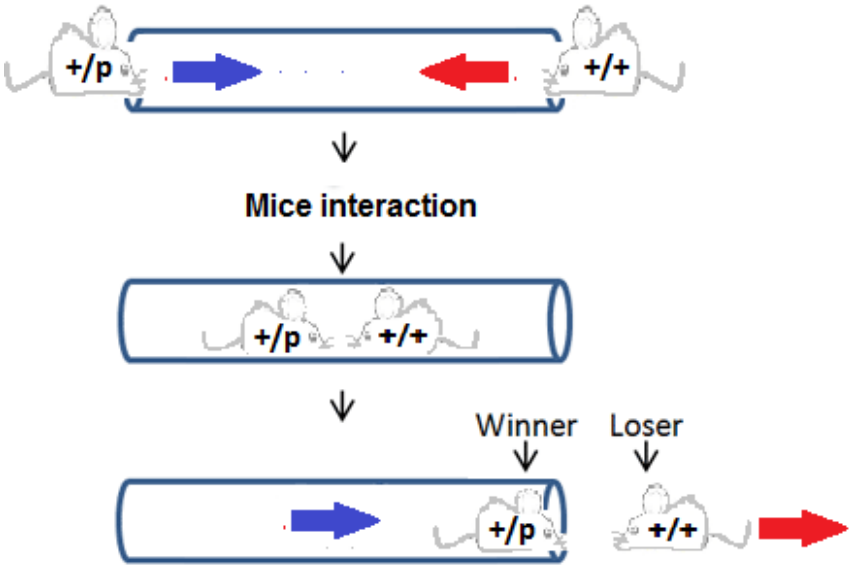


Figure 1.2. Test Tube for social dominance. In this social dominance test, males from both the wild-type (+/+) and mutant Grb10KO$^{+/p}$ mouse placed on either end of a tube, after interaction between the mice, the dominant mouse forces the other to withdraw from the tube. Withdraws mouse is the loser while the one that remains in the tube is the winner. Mutant Grb10KO$^{+/p}$ males showed to be significantly the dominant mouse (winners) compared to wild-type male (+/+) (Adapted from Hatayama et al., 2011).

Structure of Grb10 and the GRB10 pathway

GRB10 is a member of a superfamily of adapter proteins that includes Grb7, Grb10 (Figure 1.3), Grb14, and a protein of Caenorhabditis elegans called Abnormal cell migration protein 10 (Mig-10) (Morrione, 2000; Han et al., 2001). The human Grb10 gene is located on chromosome 7p12.2 with 26 exons. In the house mouse is on chromosome 11 with exon count 20 (NCBI, 2014).

Grb10 as Grb7 and Grb14 share significant sequence homology and a very conserved molecular architecture. It consists of an N-terminal region harbouring a conserved proline-rich motif (P), Ras-associating (RA) domain, pleckstrin-homology (PH) domain, Src homology 2 (SH2) domain and BPS domain which is in between the PH and SH2 domains (Figure 1.4) (Han et al., 2001; Stein et al., 2001; Holt & Siddle, 2005; Depetris et al., 2009).

The Grb10 family of adapter proteins is characterized by the presence of several splicing variants (Daly, 1998). In human, Grb10 isoforms (Figure 1.5) vary in length from 536 to 594 amino acids, there are three main Grb10 splice variants, hGrb10α (previously referred to as GRB-IR) which express 548 amino acids protein, hGrb10β (or GRB-IRSV1) which express 536 amino acids protein and it is the most widely and abundantly expressed isoform, and hGrb10ɣ which express 594 amino acids protein (Morrione, 2003; Kabir & Kazi, 2014), although, other variants of Grb10 also have been reported (Figure 1.5) which are hGrb10δ which has an extended 3′ untranslated region, hGrb10ε, and hGrb10ζ (Han et al., 2001; Morrione, 2003). Moreover, there are 26 Grb10 isoforms represented as the Ensembl transcript identifiers (Table 1.1). (Ensembl.org, 2017). In mouse Grb10 protein contains 621 amino acids (Kabir & Kazi, 2014) and it has two main Grb10 splice variants (Figure 1.5) mGrb10α (Ooi et al., 1995) and mGrb10δ (Laviola et al., 1997), mGrb10δ is 75 bp shorter, and it is predominant isoform in mouse tissues, The mGrb10α and mGrb10δ isoforms have additional mouse-specific regions (M) between the proline-rich and RA domains (Holt & Siddle, 2005). In addition, two isoforms are predicted mGrb10β1 and mGrb10β2 both differ in the alternative splicing of exons 3 and 4 (Arnaud et al., 2003). There are 9 Grb10 isoforms represented as the Ensembl transcript identifiers (Table 1.1) (Ensembl.org, 2017).

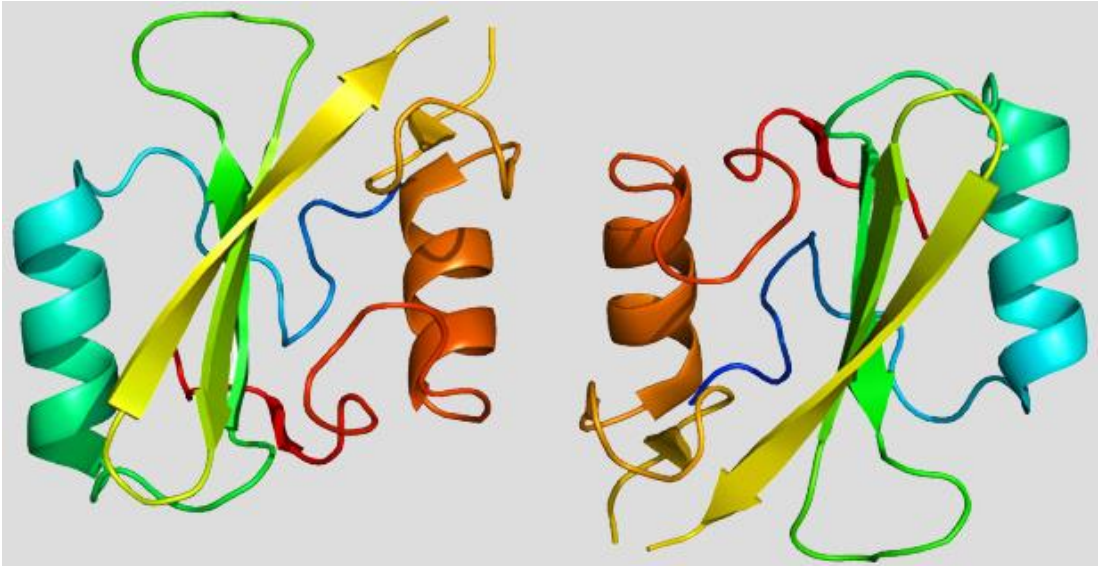


Figure 1.3. GRB10 protein. Structure of the Grb10 protein (adapted from Wikimedia Commons contributors, 2015).

Figure 1.4. Grb10 protein domains. A schematic of the five major protein domains of Grb10: the proline-rich (P) region, the Ras-associating domain (RA), the pleckstrin homology domain (PH), the Src homology 2 domain (SH2), and the intervening domain (BPS).

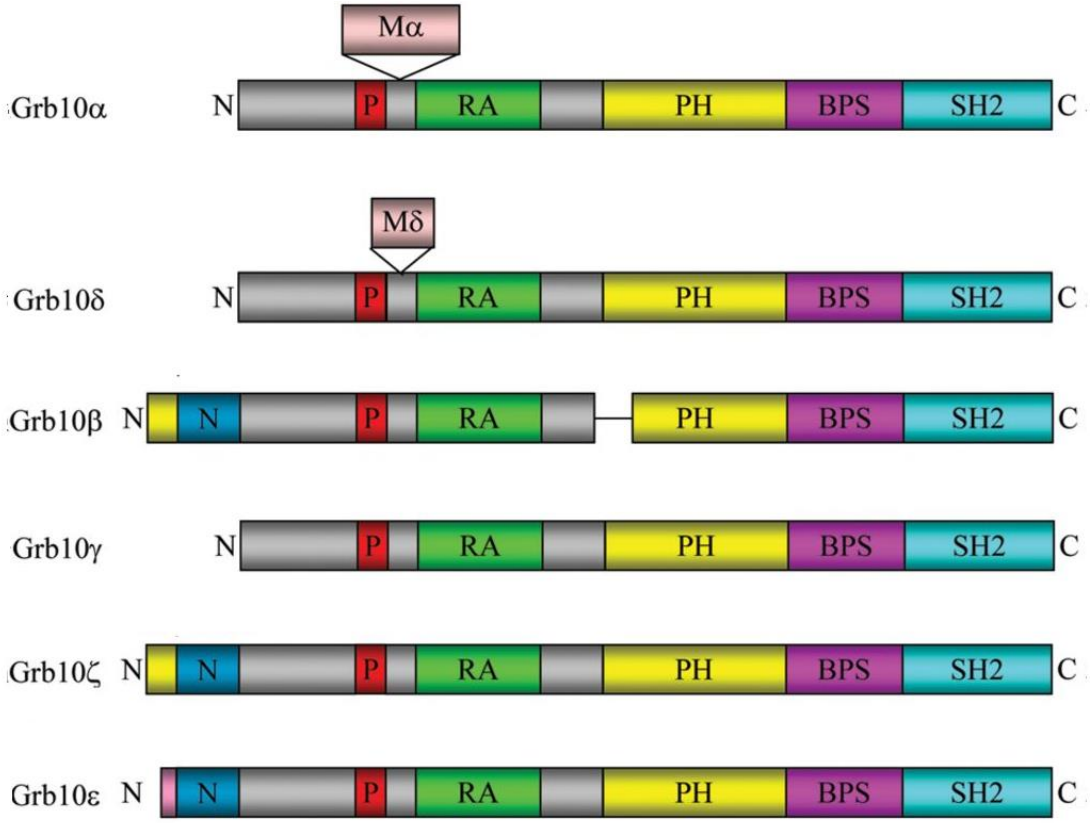


Figure 1.5. Grb10 isoforms. A schematic representation of the mouse and human isoforms of Grb10. The mouse Grb10α and Grb10δ isoforms have additional mouse-specific regions (M) between the proline-rich and RA domains. The human Grb10β, Grb10ζ and Grb10ε isoforms have a common extended N-terminal region with additional unique segments at the extreme N-terminus (adapted from Holt & Siddle, 2005).

Table 1.1. Grb10 isoforms in the mouse and human represented as the Ensembl transcript identifiers.

Name	Transcript ID	bp	Protein
mGrb10-201	ENSMUST00000093321	5061	596aa
mGrb10-202	ENSMUST00000109653	2048	550aa
mGrb10-203	ENSMUST00000109654	4755	541aa
mGrb10-204	ENSMUST00000124587	479	No protein
mGrb10-205	ENSMUST00000142877	3341	No protein
mGrb10-206	ENSMUST00000143386	761	171aa
mGrb10-207	ENSMUST00000143915	430	84aa
mGrb10-208	ENSMUST00000148254	351	No protein
mGrb10-209	ENSMUST00000150972	940	116aa
hGRB10-201	ENST00000335866	5032	536aa
hGRB10-202	ENST00000357271	2050	548aa
hGRB10-203	ENST00000398810	4785	536aa
hGRB10-204	ENST00000398812	4705	594aa
hGRB10-205	ENST00000401949	2627	594aa
hGRB10-206	ENST00000402497	2319	536aa
hGRB10-207	ENST00000402578	4760	536aa
hGRB10-208	ENST00000403097	4895	536aa
hGRB10-209	ENST00000406641	2700	536aa
hGRB10-210	ENST00000407526	2299	536aa
hGRB10-211	ENST00000428711	484	108aa
hGRB10-212	ENST00000439044	743	55aa
hGRB10-213	ENST00000461886	624	No protein
hGRB10-214	ENST00000465602	381	No protein
hGRB10-215	ENST00000467386	524	No protein
hGRB10-216	ENST00000470992	544	No protein
hGRB10-217	ENST00000473696	1065	No protein
hGRB10-218	ENST00000482397	575	No protein
hGRB10-219	ENST00000483819	657	No protein
hGRB10-220	ENST00000490051	578	No protein
hGRB10-221	ENST00000492265	447	No protein
hGRB10-222	ENST00000643299	2490	536aa
hGRB10-223	ENST00000644716	477	No protein
hGRB10-224	ENST00000644769	5431	94aa
hGRB10-225	ENST00000644879	3888	632aa
hGRB10-226	ENST00000645075	2156	588aa

The Grb10 pathway is not well understood. Studies examining the function of the maternally expressed Grb10 allele suggest that the Grb10 protein product interacts directly with a receptor tyrosine kinases and other signalling molecules that bind with insulin receptor and insulin-like growth factor receptor driving changes in downstream

gene expression (Charalambous et al., 2003; Garfield et al., 2011; Mokbel et al., 2014; Morrione, 2003). In general, receptor tyrosine kinases found in the cell in the form of separated inactivated monomers. Binding of signal molecules, which are often growth factors from the extracellular matrix, with two nearby tyrosine kinase receptor monomers, leads to activation and alteration into a dimer form (binding of two monomers). This form, in turn, promotes the phosphorylation of the tyrosine suspended in the cytoplasm converting the ATP into ADP. The phosphorylated active receptor then recognised by multiple relay proteins including Grb10 that binds to the tyrosine kinase dimer.

Overexpression of Grb10 suppresses Wnt signalling pathways, which are well established regulators of growth, by interacting with its intracellular portions at Low-density lipoprotein receptor-related protein 6 (LRP6) and interfering with the binding of AXIN1 to LRP6, which affect its normal function of inhibiting the degradation of β-catenin which regulates the coordination of cell adhesion and gene transcription (Tezuka et al., 2007).

Grb10 also has a synergistic effect on certain genes such as AKT (Jahn et al., 2002) which encoded for serine/threonine-specific protein kinase that plays a role in various cellular processes such as glucose metabolism, cell migration, proliferation, apoptosis and transcription (NCBI, 2018). Ret receptor protein tyrosine kinase which is implicated in the development of the nervous, renal and endocrine systems also interact with Grb10 SH2 as a signalling intermediate (Pandey et al., 1995). Grb10 is also considered as a positive regulator of the Kinase Insert Domain Receptor (KDR)-A Type III Receptor Tyrosine- and Vascular endothelial growth factor (VEGFR-2) signalling pathway which is an important regulator of vasculogenesis and angiogenesis (Giorgetti-Peraldi et al., 2001) as it inhibits neural precursor cell expressed developmentally down-regulated protein 4 (NEDD4)-mediated degradation of KDR/VEGFR-2 (Murdaca et al., 2004). It also binds to NEDD4 forming a complex of the NEDD4 C2-GRB10 SH2 leading to IGF1R ubiquitination by NEDD4 (Figure 1.6) (Vecchione et al., 2003; Huang & Szebenyi, 2010). Other non-receptor tyrosine kinases that Grb10 interact are Tec causing a suppression effect (Mano et al., 1998) and the Janus kinase 2 (Jak2) causing a downstream effect (Moutoussamy et al., 1998).

Grb10 SH2 domain can interact with Raf1 and MEK1 in a phosphotyrosine-

independent manner and MEK1 required this association to induce survival of the HTC-IRand COS-7 cells (Nantel et al., 1998). This domain is also associated with activated Hepatocyte Growth Factor Receptor (HGFR) and fibroblast growth factor receptors (FGFR) (Wang et al., 1999) and with receptor tyrosine kinase ELK in vascular endothelial cells (Stein et al., 1996).

Furthermore, it plays an important role in translocation of AKT to the cell membrane forming a constitutive complex (Jahn et al., 2002) although it delays and reduces AKT1 phosphorylation in response to insulin stimulation (Wick et al., 2003). Phosphorylation by mTORC1 regulates Grb10 establishing a feedback pathway by which mTORC1 inhibits the gene of INSR-dependent signalling (Hsu et al., 2011). The BPS domain of Grb10 directly inhibits substrate phosphorylation of the activated receptors tyrosine kinases (Stein et al., 2001) that leads to reduced IR protein levels and increased receptor degradation (Ramos et al., 2006). Thus the inhibitory mechanism of Grb10 mainly can be achieved by interference with the signalling pathway and increased receptor degradation, as Grb10 blocks affiliation between INSR and insulin receptor substance1 (IRS1)/ insulin receptor substance2 (IRS2) and prevents insulin-stimulated IRS1/IRS2 tyrosine phosphorylation stimulation (Wick et al., 2003).

Grb10 is also known to interact with several oncogenes in which a gene in certain circumstances can transform a cell into a tumour cell. Kazi et al (2013) stated that Grb10 associated with ligand FLT3 aside from oncogenic FLT3 and recruits p85 resulting in activation of AKT and STAT5 in hematopoietic cells. In a phosphotyrosine-dependent manner, Grb10 interact with a complex of breakpoint cluster region protein which is a complex of tumour cell (Bcr)-Abl (Bai et al., 1998), The N-terminal of Grb10 associated with Platelet-derived growth factor receptor (PDGFR), epidermal growth factor receptor (EGFR) and Abelson murine leukaemia viral oncogene homolog (c-Abl) (Deng et al., 2008), as it has been shown that it positively regulates PDGF, IGF-1 and insulin signalling, enhancing cell proliferation (Frantz et al., 1997).

Most of the analyses of potential molecular interactors of Grb10 have been carried out on tissues with a maternal expression of Grb10. The molecular interactors of Grb10 in the brain where the paternal copy is expressed and which has been associated with growth and social dominance remain unknown. Although lots of its mechanism involved large numbers of gene expression pathway.

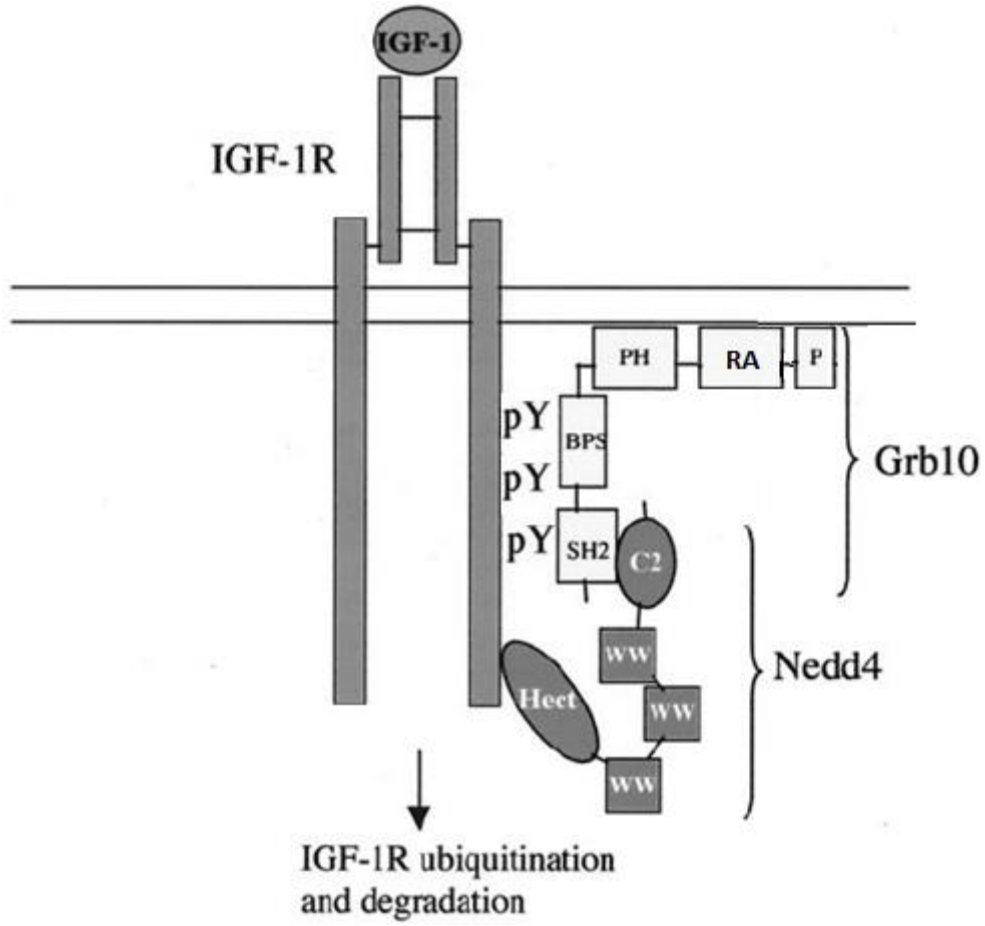


Figure 1.6. Model for Grb10 as an adapter connecting Nedd4 to the IGF-1R. The Grb10 SH2 domain constitutively binds the C2 domain of Nedd4. Upon ligand stimulation, Grb10 binds the activated IGF-IR through the BPS domain, allowing the formation of a complex that includes the activated IGF-IR, Grb10, and Nedd4. These interactions lead to ligand-mediated and Nedd4-dependent ubiquitination and degradation of the IGF-IR (Adapted from Vecchione et al., 2003).

In conclusion, Grb10 is imprinted in a highly isoform- and tissue-specific manner. It involved in several pathways and played different roles in them. Some of its functions cooperate with other proteins, some of the functions could be acted by Grb10 itself. It conserved molecular architecture which are a proline-rich N-terminal region, a Ras-associating-like domain, a pleckstrin homology domain, a family-specific BPS region, and a conserved C-terminal Src homology 2 (SH2) domain (Figure 1.4).

The mechanisms of the negative regulation of IR and IGF1R by Grb10 is not yet clear. But the biochemical studies have shown that Grb10 can bind IR and IGF1R through

their BPS and SH2 domains that have the ability to recognize phosphotyrosine-containing peptides on a variety of activated tyrosine kinase receptors (He et al., 1998; Stein et al., 2001; Depetris et al., 2005), this allows recruitment of IRSs, which, in turn, triggers activation of signalling cascades to induce metabolic, mitogenic and other cellular responses.

Aim

Here I aim to shed light on the potential molecular interactors of the paternal Grb10 allele in the brain during development and adult stages. For this, I dissected midbrain and neocortex samples from wild-type and paternal Grb10 knockout mice at five developmental stages (E18.5, 1 week old, 1 month old, 3 months old and 6 months old mice). Transcriptome profiling using Illumina next generation mRNA sequencing was then done on these samples. From this data, after curating raw read sequence data and transcriptome annotation, I carried out differential gene expression analysis to uncover sets of genes and or pathways mediating the phenotypic differences in social dominance between wild-type and paternal Grb10 knockout. Additional transcriptome data from liver where the maternal Grb10 allele is expressed was used to confirm that differentially expressed genes identified are not systemic effects of Grb10 deletion. Gene co-expression networks were then constructed to further identify additional potential Grb10 interactors in the brain. The results obtained from this project provide further insights into the genetic basis of behavioural characteristic and may also provide valuable information for the understanding of mental disorders.

Chapter II. Experimental protocols and sample collection

Introduction

In order to identify potential gene interactors of Grb10 in the developing and adult brain, I collected tissue samples from the midbrain and the neocortex. Grb10 paternal allele is expressed in the midbrain whereas the maternal copy is only marginally expressed in some small brain areas (Garfield et al., 2011). The neocortex has not been shown to express either allele. In this chapter I describe the methods followed to corroborate the spatial gene expression pattern of Grb10, tissue dissection of the neocortex and the midbrain for wild-type and Grb10KO$^{+/p}$ mutants at five different developmental time points. I also describe the methods used for the genotyping of each mouse used in the analyses to confirm if they were mutant or wild-type.

Methods

Ethics Affirmation

The mice used in this study were bred under a United Kingdom Home Office license granted to the project 101 A awarded to Prof. Andrew Ward. Mice were sacrificed using schedule one procedures.

Mice for the experiment

A total of 30 mice were used. Of these, 15 were wild-type and 15 were knockouts. The knockout line used is that reported by Garfield et al. (2011). In brief, this line was obtained by where chimeric mice were created by insertion of a *LacZ*: neomycin2 gene-trap (β-geo) cassette within Grb10 exon 8 (Figure 2.1) in embryonic stem cell line (XC302; Baygenomics) into F2 (C57BL/6 with CBA) strain blastocysts, using Nagy et al., (2003) standard methods. The microinjection of gene-trap supplanted 12 base pairs (bp) of endogenous Grb10 sequence (the 3′ terminal 11 bp from exon 8 in addition to the 5′ most 1 bp from intron 8). These mice were subsequently maintained on a C57BL/6 with CBA mixed genetic background. All mice were housed under the standard conditions in accordance with UK Home Office Regulations in individually ventilated cages at a density of 1-6 mice per cage with 13 hours of light and 11 hours

of darkness including 30 minutes of artificial dawn and dusk lighting, the temperature is maintained at 21 ±2 °C with a relative humidity of 55 ±10%. They were provided with unrestricted access to water and rodent maintenance diet food (RM1 formula) that contains 20% of protein. The mice bred at the age of 2 months, Once the male has efficaciously mated the female, the semen forms hard waxy 'plug' (white colour) in the vagina, the vaginal plug will be then checked on the second day of mating, determining the success of mating and pregnancy. The gestation period ranged between 19-21 days, after that the pregnant mouse gives birth to baby mice (pups) their numbers can range from 1 to 12. Mice weaned at the age of 3 weeks after birth, the weaned mice removed from the parent cage and separated by sex. Post weaning mice were housed together in the same cage according to their age group.

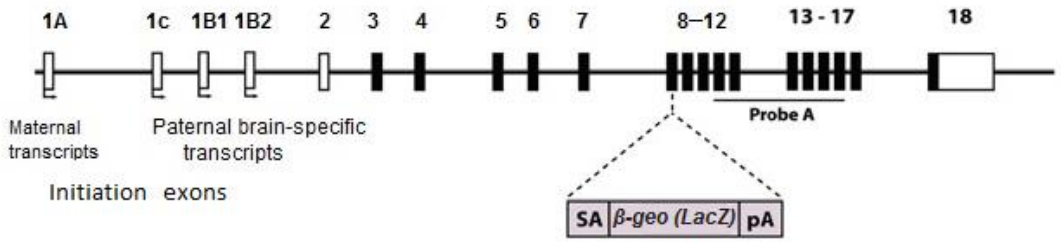


Figure 2.1. The structural features of mutant Grb10. Mouse Grb10 locus is showing the *LacZ*: neomycin[2] (β-geo) insertion in the Grb10 allele, transcriptional start sites (arrows), numbered exons (boxes) and translated regions (filled boxes). The integrated gene-trap cassette includes splice acceptor (SA) and polyadenylation (pA) signals, and a *LacZ* reporter (Adapted from Garfield et al., 2011).

Stain base for β-galactosidase (*LacZ*) assay

β-galactosidase is a glycoside hydrolase enzyme that catalyzes the hydrolysis of β-galactosides (include carbohydrates containing galactose where the glycosidic bond lies above the galactose molecule) into monosaccharides through the breaking of a glycosidic bond. To assess the spatial patterns of Grb10, an assay to reveal *LacZ* neomycin2 gene-trap cassette which had been inserted within Grb10 exon 8 (Figure 2.1) was used.

For this, the dissected tissue samples were fixed in 4% paraformaldehyde in phosphate-buffered saline (PBS) (pH 7.0-7.5) for 2 hours at room temperature, and then submerged in X-gal staining solution and incubated in the dark at 28 °C

overnight. X-gal used to detect the presence of β-galactosidase (the *lacZ* gene) which forms a blue product after cleavage by β-galactosidase that can be used as a reporter in various application.

Developmental stages sampling

For this project, a total of 30 male mice were used which were grouped after genotyping into 15 wild-types (WT) mice and 15 paternal mutant Grb10KO$^{+/p}$ mice.
5 different developmental age groups have been selected Embryo 18.5 (after Grb10 gene expression is established), 1 week old (before weaning), 1 month old (after weaning and juvenile stage), 3 months old (sexually mature stage) and 6 months old mice (mature adult stage) to find out if differences in gene expression in the paternal knockout brain remain constant throughout development. From each mouse, the subcortical areas and the cortex tissue samples were obtained that gives a total of 60 tissue samples. To comprehend the progressions at transcription levels at the different patterns of variation within across at the certain brain regions in WT and mutant Grb10KO$^{+/p}$ mice, a total of six mice (3 WT and 3 mutant Grb10KO$^{+/p}$) were sacrificed and the brain sample was dissected at designated time points which are: Embryo 18.5 (E18.5), 1 week (1W), 1 month (1M), 3 months old (3M) and 6 months old (6M). Not all the replicates at different age groups are littermate (Supplementary Table S2.3).

Brain dissection

Before and throughout the dissection, details of the mouse were recorded (Supplementary Appendix S2.1, S2.2), including their parent information such as their date of birth and numbers (Supplementary Table S2.3). All the set for dissection and samples collection were prepared before the further procedure (Figure 2.2). Immediately before dissection, each mouse was sacrificed by cervical dislocation, sprayed with 70% of ethanol to disinfect the area and control spreading of hairs in brain samples, and then decapitated. For dissection of mouse embryos at E18.5 stages, previously sacrificed pregnant female mice were sprayed with 70% of ethanol on the abdominal area. Using a scalpel access to the uterine sack was achieved. E18.5 embryos were removed and kept in an ice-covered petri dish that containing

PBS solution. Embryos to be dissected were then decapitated.

For mice of all ages, the skull was dissected to take out the whole-brain (Figure 2.3A). The brain tissue sample then kept in a petri dish, which then longitudinally bisected (Figure 2.3B). The cerebellum was then removed and dissection of the neocortex from subcortical areas was carried out.

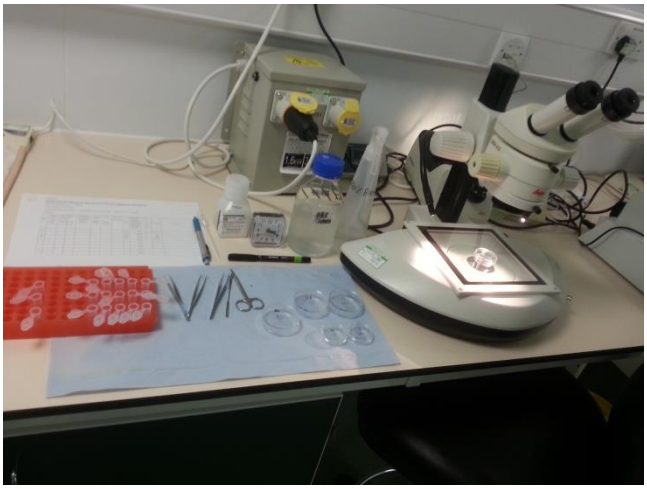


Figure 2.2. Preparation before dissection. Pre-dissection set preparation.

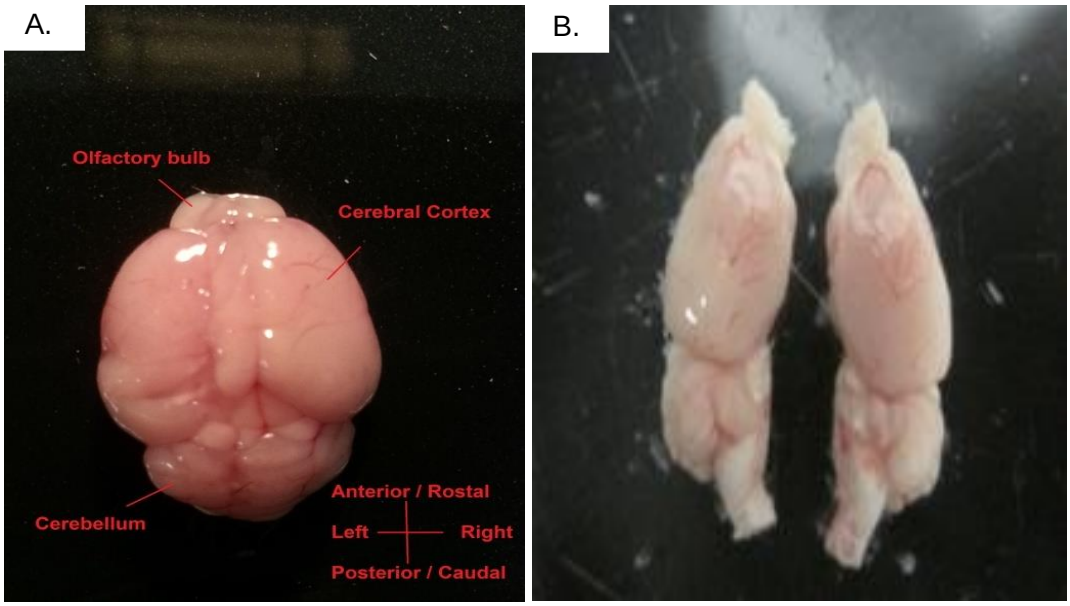


Figure 2.3. Brain schematic of Adult Mouse. A, brain tissue sample of 3 months old mouse. B, longitudinally bisection of the brain tissue sample.

The dissecting magnifying lens stereomicroscope (LEICA MS5 Microsystems, Bensheim, Germany; Figure 2.2) was used in order to accurately separate the neocortex from subcortical areas (Figure 2.4).

The neocortex and the subcortical areas of each hemisphere of the brain tissue sample were kept in a separate Eppendorf tube filled with RNAlater to maintain RNA integrity during storage of the tissue samples. The samples then kept at 4 °C overnight to allow the RNAlater thoroughly penetrate the tissue, after that the samples were stored at -80 °C for long-term storage until they are used for further processing. The tail snip of each mouse was kept inside an empty Eppendorf tube for future investigation of genotyping if it required distinguishing between the WT and mutant Grb10KO$^{+/p}$ mice.

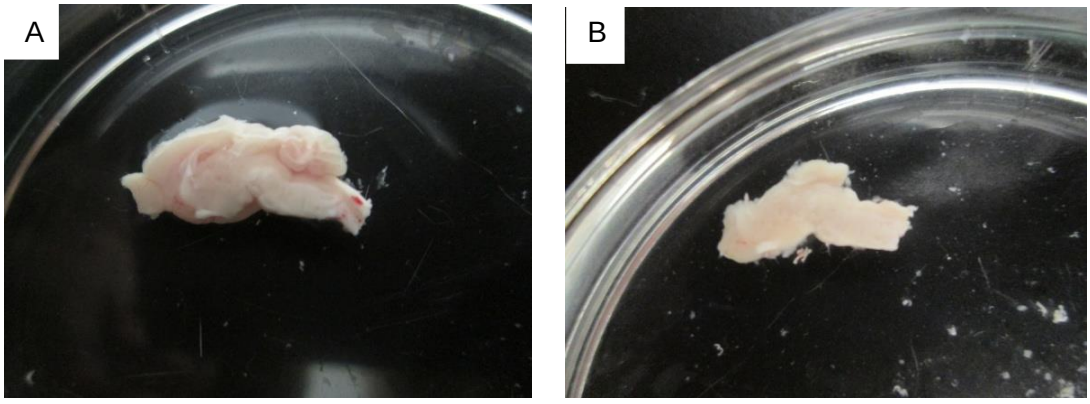


Figure 2.4. The right hemisphere of the brain tissue sample. A, side view of the right hemisphere of the brain tissue sample. B, subcortical areas of the brain tissue sample (other regions such as the cortex, olfactory bulb and cerebellum were removed).

To prevent RNA degradation of the brain tissue samples which may impact on gene expression profiling (Optiz et al., 2010) the time interval between sacrificing of the mouse and storing its brain tissue sample in RNA later solution was recorded. Only samples which were secured in less than 5 minutes were used.

In this study, only samples from male individuals were used. Female pups at age of E18.5 and 1W of age were identified by checking the genital organ (uterine horn). while for other age groups of mice (1M, 3M & 6M old) determination of the gap between the anus and the genital organ (anogenital distance) was carried out, in which in males is approximately two-fold that of the female.

Sample storage, RNA extraction and sequencing

To reveal the molecular pathways and or sets of genes interceding ability of Grb10's to regulate social dominance between WT and mutant Grb10KO$^{+/p}$ mice and to investigate differentially expressed genes during developmental stages in the cortex and subcortical areas of the brain. RNA-Seq is required which is an accurate measurement of gene expression at the transcript level, it gives high dynamic diversity and almost unbiased view of the transcriptome landscape (Mortazavi et al., 2008, Sultan et al., 2008). It generates both qualitative and quantitative data providing visibility to undetected changes occurring in mutant Grb10KO$^{+/p}$ as compared to the WT. For this purpose, the whole cortex and the subcortical areas of the right hemisphere of the brain samples were sent to BGI Tech Solutions (Hong Kong) Co., Ltd for RNA extraction, library preparation and sequencing transcriptome. Illumina-HiSeq 2500/4000 was performed on each of the 60 samples for sequencing (paired-end reads). The following requirements for the samples were considered during RNA extraction:

1) Sample Type: Integrated total RNA samples that are treated with DNase. Avoid protein contamination during RNA isolation.
2) Sample Quantity: total RNA≥ 1μg.
3) Sample Concentration: concentration ≥ 80ng/ul.
4) Sample Purity: OD260/280 ≥ 1.8, OD260/230 ≥1.8; RIN ≥ 7.0.

The final results of raw RNA-Seq datasets were received from the BGI Tech Solutions (Hong Kong) Co., Ltd in 500 gigabytes (GB) hard disk.

Genotyping

At three weeks of age, mice were identified by punching their ears using a special ear punch tool (INS750075-5) to produce a small (0.5 to 2 mm) notch around the edge circumference of the ear (Figure 2.5). The tissue sample of the ear snip was genotyped immediately or stored at -20 °C for genotyping. A biopsy of tail snip was used for E18.5 and 1 week old mice.

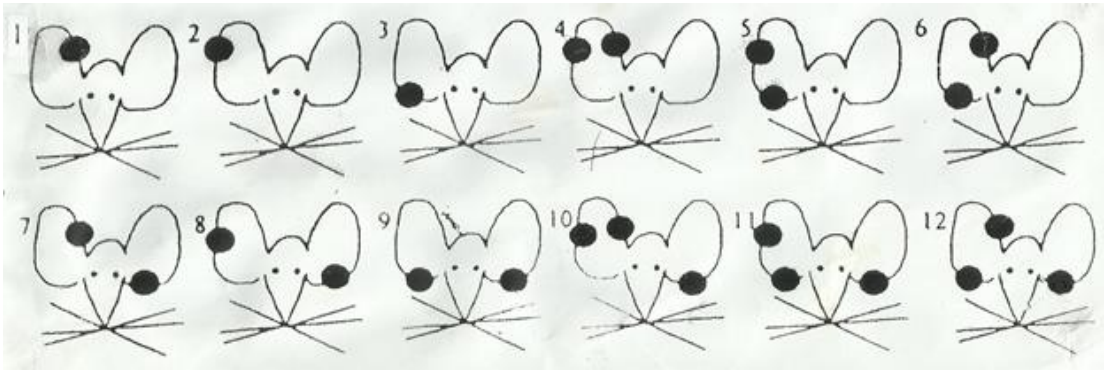


Figure 2.5. Mice identification. Ear punching is used to mark individual mouse using a special punch tool to produce a small notch around the edge circumference of the ear.

PCR was used to identify the genotype of the WT and the mutant Grb10KO$^{+/p}$ mice at the 5 different age groups. It is a technique used to make multiple copies of a segment DNA of interest generating a massive range of copies from a small initial specimen. Animals were genotyped by PCR using Forward and Reverse primers (Table 2.1) specific to the Grb10 these primers amplify the parental Grb10 allele, which is present in both WT and mutant tissue samples. In addition to the KM4 reverse primer (Table 2.1) specific to the mutant allele which amplifies only the mutant allele in mutant tissue samples.

Table 2.1. The sequences of primers used in PCR. Animals were genotyped by PCR using forward and reverse primers specific to the Grb10 and KM4 reverse primer specific to the mutant Grb10.

Primer	**Sequence**
KM4 reverse primer	5′-CACAACGGGTTCTTCTGTTAGTCC-3′
Forward primer	5′-CCAAGTGGAGAGTACCATGCC-3′
Reverse primer	5′-TCACCTGACAGGCACCTCCCC-3′

To carry out the PCR, 600 microliter (µl) of freshly made of 100 millimolar (mM) sodium hydroxide (NaOH) was added to each biopsy sample (tail snip or ear clip) in a 1.8 ml eppendorf tube, all the tubes placed in the test tube rack, the lid of the eppendorf tube then pierced with a needle to avoid the pressure during boiling, the test tube samples were boiled for 10 minutes, and then cold down at room temperature. 100 µl of 1M Tris-HCl (pH 8) was added to each sample to neutralise the sample. 14 µl of a master mix sample was made which consist of 7.5 µl of 2x PCRBio Red (containing Taq DNA

polymerase, dNTPs, buffer and loading red dye), 0.74 µl of the primers, and 5.75 µl of Milli-Q (MQ) water. In new PCR tubes 1µl of each target fragment and 14 µl of master mix were added and mixed, then ran into PCR machine (thermocycler) that carries out the temperature cycles, with the following settings: an initial denaturing set up at 94 °C for 9 min followed by 32 cycles (1 min at 94 °C, 1 min at 62 °C, and 1.5 min at 68 °C) and a last step at 68 °C for 8 min. The following reaction occurred during running the PCR, first the temperature is raised to near boiling (94 °C), causing the double-stranded DNA to denature (separate) into single strands, when the temperature is decreased (62 °C), short DNA sequences (primers) anneal to the complementary matches on the sequences of the target DNA. The primers bracket the target sequence to be copied. At a slightly higher temperature (68 °C), the enzyme taq polymerase binds to the primer sequences and adds nucleotides to extend the second strand, where the first cycle is completed. In subsequent cycles, the process of denaturing, annealing and extending are continually making additional DNA copies. After few cycles, the target sequence defined by the primers begins to accumulate. At the end of the 32 cycles, many duplicates of the target sequence produced from a single starting molecule.

During running thermocycler, a gel was prepared that contains 1% of agarose powder (0.7 gm), 99% of Tris-Acetate-EDTA (TAE) solution (70 ml), and 0.5 µg/ml of Ethidium Bromide solution (3.5 µl). The wells of the gel were made via placing a comb into the slots in the tray. In the first lane, 6 µl of the ladder solution (500 bp) was added and 15 µl of each sample was added to the remaining lanes. Subsequently, electrophoresis machine ran on the gel at 110 volts for 35 minutes. Eventually, the gel ran under UV-Transilluminator machine, and the result was saved and recorded.

Results

Allele-specific Grb10 gene expression in the brain and heart

To assess tissue-specific expression patterns of Grb10 paternal and maternal alleles a *LacZ* assay was carried out to track *LacZ* the neomycin2 gene-trap cassette inserted in the knockout mice (Garfield et al., 2011). The test was carried out on dissected brain and heart samples of mutant Grb10KO$^{m/+}$ and mutant Grb10KO$^{+/p}$ on 5 weeks old mice, and Grb10KO$^{m/+}$ on 1 day old pup. This test was also carried out to validate

the brain regions to be analysed and to ensure brain dissection protocols avoided contamination between brain areas expressing Grb10 and areas not expressing Grb10. The maternal copy of Grb10 was found to be expressed in the heart (Figure 2.6) whilst the paternal copy is clearly expressed in the midbrain area (Figure 2.7). The neocortex shows little evidence of the *LacZ* probe (Figure 2.7).

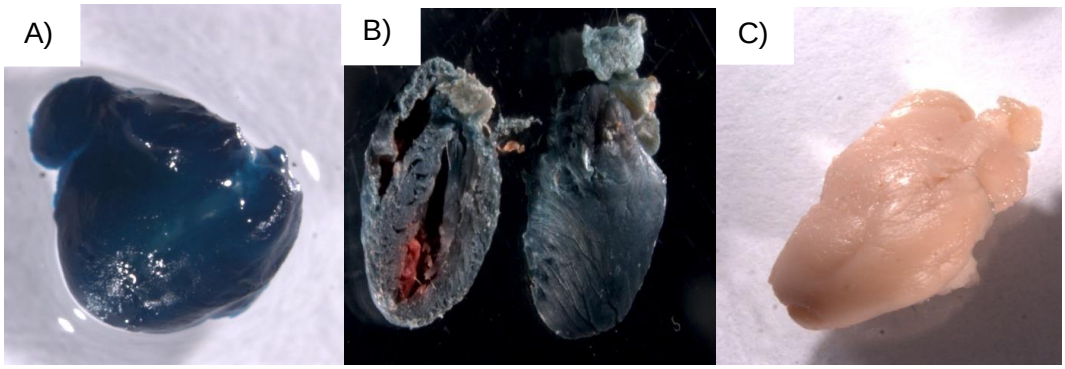


Figure 2.6. *LacZ* reporter Grb10 expression in the brain and heart. Maternal mutant Grb10 showing extensive expression in the heart of A) one day old pup and B) five weeks old mice. C) Paternal mutant Grb10 shows no Grb10 expression in the heart of five weeks old mice.

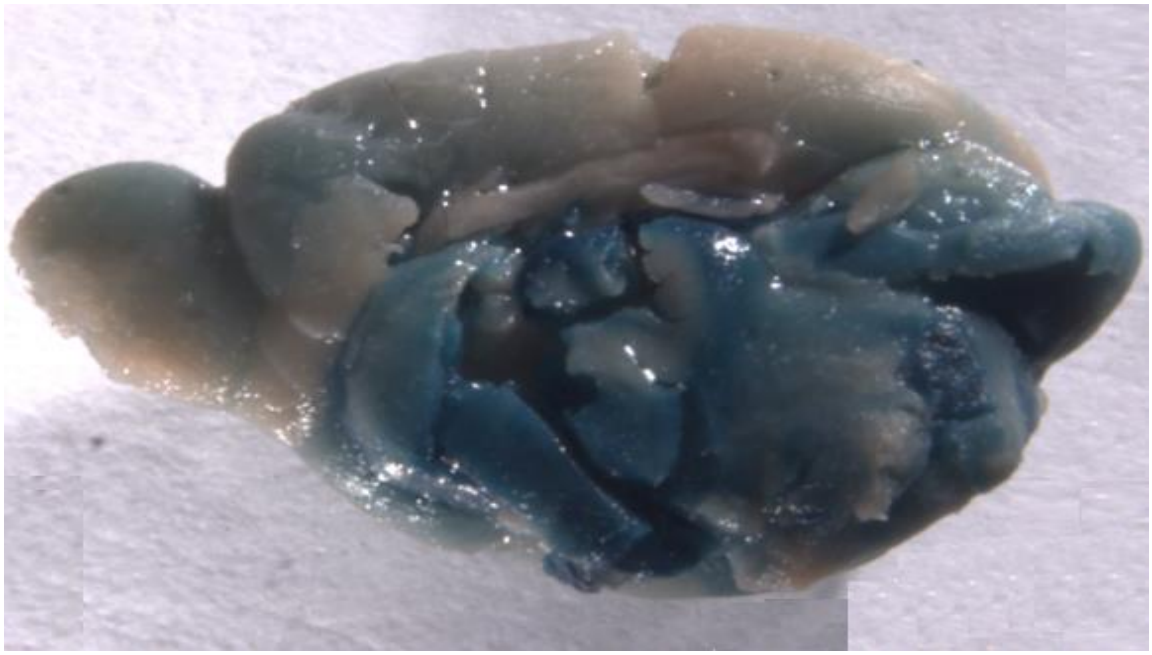


Figure 2.7. Brain sample of a mutant Grb10KO$^{+/p}$ mouse. *LacZ* reporter expression in specific regions in the brain of the paternal Grb10 mutant mouse, *the cerebellum was detached.

In this test, the *LacZ* reporter revealed high expression in the heart of the maternal Grb10KO while there was no expression in the heart of the paternal Grb10KO$^{+/p}$ (Figure 2.6), on the other hand, *LacZ* reporter expression in the brain showed no expression from the maternal Grb10KO while in paternal Grb10KO$^{+/p}$ the expression was clear in specific regions of the brain (Figure 2.7). Both results confirmed the expression profile identified previously by Garfield et al. (2011) and Cowley et al. (2014).

Sample collection

Having established that the paternal Grb10 allele is indeed expressed in the subcortical areas but not in the neocortex, I proceeded to collect samples for three wild-types and three paternal KO mice at five different developmental time points that had mentioned earlier. Sampling over several time points allowed me to be able to identify a broader range of potential interactors of Grb10 in the brain across development. Because of resource constraints I only used male subjects for this study. Samples were collected in a timely fashion immediately after sacrificing each mouse to minimise RNA degradation. In order to minimise other genetic variation to impact on my results. For this study, PCR genotyping was used to distinguish between the mutant Grb10KO$^{+/p}$ mice from the WT mice. The agarose gel electrophoresis which is a diagnostic tool used to visualise the fragments revealed double bands for the mutant Grb10KO$^{+/p}$ mice and a single band for WT mice as shown in Figure 2.8.

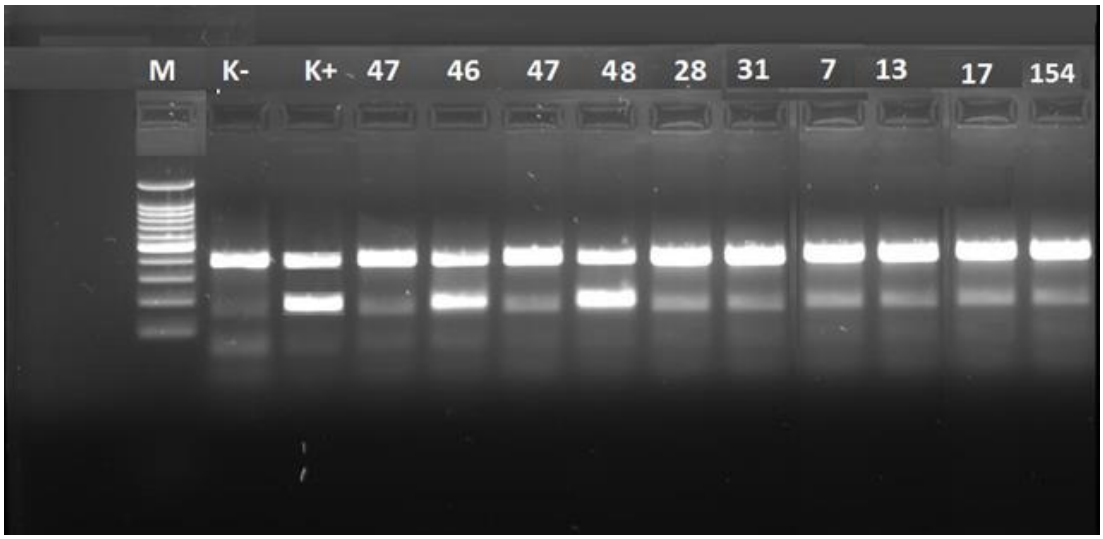


Figure 2.8. Detection of mutant Grb10KO$^{+/p}$ samples using PCR. Agarose gel electrophoresis showing Lane M: 500-bp DNA molecular weight markers; Lane K-: negative control represents one band; Lane K+: positive control represents double bands, Lanes for sample number 46 and 48 showing positive PCR reaction for mutant Grb10KO$^{+/p}$ mice; Lanes for sample number 47, 28, 31, 7, 13, 17 and 154 showing negative PCR reaction for WT mice.

Discussion

There is a remarkable conservation of the neural circuitry involved in regulation of social behaviour across vast taxonomic distances (Goodson 2005; Newman 1999). Thus, the use of animal models is a powerful tool to gain a better understanding of the neurological and molecular underpinnings of complex social behaviour. The use of mutant rodent models has allowed the identification of genes underlying a variety of phenotypes. Grb10 is an imprinted gene in human and mice with a tissue-specific manner expression (Arnaud et al., 2003; Hikichi et al., 2003; Monk et al., 2009). In the mouse, maternal allele expressed in most of the tissue while paternal allele expressed in certain parts of the brain (Garfield et al., 2011). The development of a KO line for Grb10, an imprinted gene mainly associated with regulation of growth led to the discovery of the expression of the paternal allele in the brain which lasts into adulthood and which was then associated to social dominance (Garfield et al., 2011).

Here I confirmed the spatial patterns of Grb10 expression in the mouse brain using a *LacZ* reporter for the mutation introduced by Garfield et al. (2011). According to Garfield et al. (2011) in late embryonic development limited expression of maternal allele found in ventricular ependymal layers, epithelium of choroid plexus and

meninges of the brain. In contrast paternal allele expression is extensive throughout development in diencephalon, ventral midbrain, and medulla oblongata while in adulthood as shown in the Figures 2.7 and 2.9 the expression is within the forebrain including septum, ventral striatum, basal forebrain, thalamus, hypothalamus, midbrain, ventral tegmental area, and in hindbrain in pons and medulla excluding the cerebellum. Most of the paternal allele expressing brain areas contain a number of regions critical for the coordination and expression of social behaviour, indicating that Grb10 play roles in regulating social behaviour (Curley, 2011). Conversely, Grb10 paternal expression is absent from specific regions of the brain such as the cerebral cortex and hippocampus both which although are also crucial for social behaviour, are not part of the more primitive emotional circuitry (Curley, 2011). This result is consistent with Grb10 having evolved a role distinct from the classical imprinted pattern in growth to also acquiring a role in the regulation of behaviour in mammals. I also collected samples for transcriptome profiling to gain insights into the genetic pathways mediating the phenotype on the social dominance of Grb10. I used a PCR technique to distinguish between WT and mutant Grb10 mice. This method was initially developed in 1983 by Kary Mullis and co-workers (Saiki et al., 1985). It is used in research labs to amplify a particular section of DNA or gene generating thousands to millions of copies.

In primates (including human) social grooming plays a major role in social bonding (Dunbar, 2010). Garfield et al. (2011) revealed that social dominance with paternal Grb10 knockout mice more likely to carry out allogrooming & barbering behaviour with the cage mates.

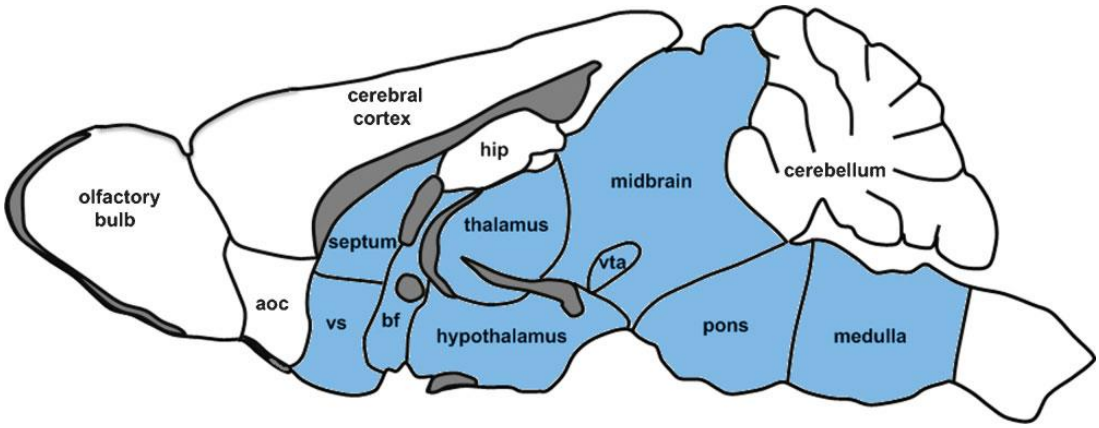


Figure 2.9. Diagram showing the paternal expression of Grb10 (blue colour) in the brain of an adult mouse. Paternal Grb10 localised at the forebrain, including the septum, ventral striatum (vs), basal forebrain (bf), thalamus, hypothalamus, midbrain, ventral tegmental area (vta), pons, and medulla including significantly nearly all monoaminergic cell populations within these areas. Cortex, hippocampus (hip) and cerebellum show no paternal Grb10 expression while slight expression observed on olfactory bulb only on nucleus olfactory tubercle and islands of Calleja which is a packed cell clusters content of olfactory tubercle (Adapted from Curley, 2011).

Chapter III. Transcriptome profile analyses of the Grb10 paternal $KO^{+/p}$

Introduction

Grb10 is a unique imprinted gene which is essential for normal embryonic development and growth. Despite the consistently reported role of Grb10 in metabolic insulin action in fat and muscle cells, its expression in the central nervous system of mice indicates its role in regulating neuronal insulin signalling and energy metabolism as well as the social dominance within the mice. The genetic interactors of Grb10 in the brain are poorly understood. Using the mouse as a model, I investigated the transcription profiles of Grb10 (mutant Grb10KO$^{+/p}$) mice compared with the wild-type mice at five different developmental time points.

Profiling of the transcriptome can be a powerful tool to establish the molecular priorities in cells at a specific point in time, condition and or tissue. Transcriptome analysis can help understanding the molecular context, gene function and phenotypical difference between individuals, developmental phases, tissues and disease states. Profiling of the transcriptome can be done mainly by two technologies microarrays and RNA sequencing (RNA-Seq). During the first decade of this century, microarrays were widely used. The disadvantage of this technique is that it measures only known genomic sequence information, as it uses probes for a predetermined set of genomic areas, sequences which are not printed in the microarray chip cannot be detected (Zhao et al., 2014).

RNA-Seq is increasingly being used for transcriptome profiling as costs of this technique have decreased. It is considered to be a more powerful technique compared to microarray technology as it detects more differentially expressed genes, and allows to identify new transcripts as reference genome annotations are revised (Zhao et al., 2014), as raw reads can be repeatedly realigned to updated reference genomes, and filtering of the reads to match unique locations of interest in the genome. Although many researchers still using microarrays platform however it is clear that in the future transcriptome analysis using RNA-Seq platform will become the predominant approach.

Transcriptome annotation provides insight into the biological process and function of transcripts at a specific point in time in a sample of one, many cells or tissues. There are two general approaches for transcriptome analysis which are guided assembly where sequence reads are mapped onto a reference genome and de novo

transcriptome assembly which using software to construe transcripts from short next generation sequencing reads without the use of a reference genome.

For species with a well-annotated genome such as the mouse or human, guided assemblies usually provide better results. Transcriptome assemblies allow identification of transcript isoforms and to quantify relative RNA abundance per transcript and per isoform in each sample. RNA-Seq of certain biological problems required slightly different computational methods, but they all relied on a set of core approaches including alignment to a transcriptome or reference genome, assembly into transcriptional units, and quantification of expression (Garber et al., 2011).

Until recently, transcriptome annotation using guided assemblies relied on the alignment of one read at a time against the reference genome. It is very common to use reference genome for aligning the sequenced reads of RNA-Seq data, reconstructing transcripts, and quantifying their expression (Trapnell et al., 2012). This process of alignment is slow causing an analytic bottleneck (Bray et al., 2016). Some more recent transcriptome annotation programmes use a more loose alignment process but without reducing the accuracy of annotation. These include software tools: RNA-skim (Zhang et al., 2014), RapMap (Srivastava et al., 2016), Sailfish (Patro et al., 2014), Salmon (Patro et al., 2017), and Kallisto (Bray et al., 2016).

Kallisto (Bray et al., 2016) is a program for quantifying abundances of transcripts from RNA-Seq data, or more generally of target sequences using high-throughput sequencing reads. It builds an index of k-mers (generated from transcripts) from a reference set of transcriptome and then estimates expression level from the exactly matched reads directly (Bray et al., 2016). It performs pseudo-alignment for rapidly determining the compatibility of reads with targets, without the need for alignment and quantity abundances transcripts of the RNA-Seq dataset at optimal duration and accuracy (Bray et al., 2016). According to Bray et al. (2016), Kallisto showed to be more accurate than different quantification programs such as Cufflinks, Sailfish, eXpress during analysis of the median relative difference in the estimated read count of each transcript and it is faster than Cufflinks, RSEM, eXpress and Sailfish. Kallisto can perform pseudo-alignment and quantity of abundances transcripts of the RNA-Seq dataset at optimal duration and accuracy (Bray et al., 2016).

Differential gene expression

In order to identify genes which are over or underexpressed in certain conditions of genotypic backgrounds, statistical analyses of differential gene expression are widely used (Tusher et al., 2001; Grant et al., 2005) This approach although often using different normalisation protocols are applicable to both microarray and RNA-Seq data. Detection of differential gene expression depends on the sequencing depth of the sample, the length of the gene, and the expression level of the gene (Oshlack & Wakefield, 2009; Bullard et al., 2010).

Many computational methods have been developed for differential gene expression analysis. The best performing tools tend to be DESeq/DESeq2 (Anders & Huber, 2010; Love et al., 2014), edgeR (Robinson et al., 2010), and limma-voom (Ritchie et al., 2015). Some of the tools can only perform the pair-wise comparison, others such as limma-voom, maSigPro DESeq, and edgeR can perform multiple comparisons at a time. Although DESeq and limma-voom tend to be more conservative than edgeR for better control of false positives, in an experiment with fewer than 12 replicates edgeR is recommended (Schurch et al., 2015). Therefore, many researchers suggested increasing the number of biological replicates than the sequencing depth of single samples for analysis of differential gene expression (Rapaport et al., 2013; Ching et al., 2014; Liu et al., 2014; Gierli´nski et al., 2015). The algorithms of all these tools have the advantage that the entire analysis can be conducted efficiently within the *R* software environment.

Methods

Grb10 and β-geo Primers

The presence of the primers specific to Grb10 that consist of the following sequences F, 5′-CTGACCTGGAAGAAAGCAGC-3′ and R, 5′-GATCCTGTGAGACTCCTCGC-3′ and the inserted β-geo gene-trap cassette (β-geoF2 5′-CCGAGCGAAAACGGTCTGCG-3′ and β-geoR2 5′-CTTCCGCTTAGTGACAACG-3′) which being reported by Garfield et al. (2011) (Figure 2.1) were checked within all the set of paired-end sequence reads for all the 60 samples at all the 5 developmental

stages in both the subcortical areas and the cortex using bash command line.

Quality Analysis using FastQC

Quality control of paired-end sequence reads of raw RNA-Seq datasets for all the 60 samples (the cortex and the subcortical areas) at all the different stages of mice was assessed using the FastQC tool (Andrews, 2010) of Galaxy version 0.67. Three parameters needed to execute this tool, '*Short read data from your current history*' was used as this specifies raw RNA-Seq dataset of each sample. The other two parameters which are '*contaminant list*' and '*Submodule and Limit specifying file*' were left as default blank.

Transcript annotation

Locations of genes and all the coding regions in each of the sample genome were obtained using Galaxy version 0.67 tools (public web access: http://usegalaxy.org) to run samples of RNA-Seq dataset. The TopHat tool was used for mapping the RNA-Seq reads to the genome, then annotated using Cufflinks tool. The replica samples (=3) of the WT and the mutant Grb10KO$^{+/p}$ for each stage independently were merged using Cuffmerge tool. For the final stage, the differentially expressed genes executed using Cuffdiff tool. The public Galaxy server has few disadvantages: depending on the number of users using public web access: http://usegalaxy.org the data processing creates unnecessary delay giving responsibility for the result after several hours or days for each sample especially while running the TopHat tool, the list for the total number of the annotated transcriptome in Cufflinks is not equal in each sample creating difficulty for annotating of the genes for any further analysis, the large datasets must be cleaned up periodically due to the limitation of the data saving and the Galaxy server doesn't give raw counts of the transcriptome it gives only Fragments Per Kilobase of transcript per Million mapped reads (FPKM) counts. Therefore, Kallisto program used in Linux operating system for annotation of the transcriptome using the labs' own servers. The latest version of the Kallisto program dated 02/06/2016 was downloaded (http://pachterlab.github.io/kallisto/download) with a mouse ENSEMBL 87 reference transcriptome (GTF file) to build an index from a fasta formatted file of target sequences. Per gene RNA abundance was obtained by adding up expression levels per transcript

for each Ensembl Gene IDs. Estimated total read counts were conducted for each sample on the raw RNA-Seq reads using bash command line.

Differential expression analysis

For the following statistical analysis, the large-scale calculation was carried out in *R* software (version 3.2.5 (2016-04-14) -- "Very, Very Secure Dishes"). Differentially expressed genes between the wildtype and the mutant Grb10KO$^{+/p}$ samples were identified using the edgeR package supported in R (Robinson et al., 2010). The trimmed mean of M-values (TMM) normalisation method proposed by Robinson and Oshlack (2010) was performed which is displayed by the normalisation factors (calcNormFactors) in edgeR (Robinson et al., 2010). The normalised expression data on each developmental stage for each replica n =3 was analysed and the up-regulated and down-regulated differentially expressed genes with false discovery rate (FDR) < 0.05 were quantified.

Multidimensional scaling plots (MDS) and Heatmap

To explore the expression similarity between the tissue samples within the five developmental stages, MDS analysis was performed between the subcortical areas and the cortex on the normalised data samples of WT mice and mutant Grb10KO$^{+/p}$ mice separately. It is also used to discriminate between the subcortical areas and the cortex (Supplementary Figure S3.3). Heatmap package endured in the *R* software proposed by Ploner A. (2015) was performed to depict the graphical representation of the up-regulated and down-regulated differentially expressed genes datasets at each stage. Bar graph was used to represent the actual size of up-regulated and down-regulated differentially expressed genes for the five developmental stages represented in the subcortical areas and the cortex.

Functional enrichment analysis

To get further insight into the function and the correlation of the differentially expressed genes among each other’s, a potential enrichment of specific functional categories

among the up- and down-regulated genes gene ontology (GO) enrichment analysis had performed. Gene ontology annotations cover three sets that include a set of biological processes, a set of cellular components and finally a set of molecular functions. For this project, the first two sets have been used. For each tissue, the sum of all the up- and down-regulated differentially expressed genes across all five developmental stages had used. GO analysis had conducted in each of the four modules of genes; namely: up- and down-regulated differentially expressed genes in the subcortical areas and up- and down-regulated differentially expressed genes in the cortex. The lists of genes annotated to these two sets which obtained from GO downloaded from Ensembl Biomart database version 87 (http://www.ensembl.org/biomart). Separate GO enrichment analysis for each group of ontology terms had performed on each of the modules detected. Statistical significance had numerically assessed by counting within each module the number of genes annotated to each GO term and compared these counts to those derived from 10,000 equally sized random samples from the entire gene population (Monte Carlo simulation) to specify a Z-score with their corresponding p-value, finally the p-values were adjusted using for multiple testing using Benjamini-Hochberg correction.

Results

Inserted cassette is only detected in Grb10KO$^{+/p}$ tissue samples

Checking the presence of primers specific to Grb10 (probe A, Figure 2.1) (F, 5′-CTGACCTGGAAGAAAGCAGC-3′ and R, 5′-GATCCTGTGAGACTCCTCGC-3′) and the inserted β-geo gene-trap cassette (β-geoF2 5′-CCGAGCGAAAACGGTCTGCG-3′ and β-geR2 5′-CTTCCGCTTAGTGACAACG-3′) within all the set of paired-end sequence reads for all the 60 samples at the 5 developmental stages (E18.5, 1W, 1M, 3M and 6M old) in both the subcortical areas and the cortex using the bash command line revealed that in case of primers specific to the β-geo insertion WT mice as it should, had shown no presence of these primers at all the developmental stages in both the subcortical areas (Table 3.1) and the cortex (Table 3.2), in contrast mutant Grb10KO$^{+/p}$ mice had shown the presence of only forward primer at all the stages in both tissues (Table 3.3 and Table 3.4), although in the subcortical areas (Table 3.3) the total number has been nearly double that in the cortex (Table 3.4).

Table 3.1. Primers specific to β-geo insertion and Grb10 in the subcortical areas in wild-type mice. Reports for detecting primers specific for inserted β-geo gene trap cassette and the Grb10 in the subcortical areas (Subcor) at all the 5 developmental stages in (=3 at each stage) on the set of paired-end reads (F= forward reads, R= reverse reads) wild-type mice using bash command line.

wild-type Subcortical areas at E18.5						
Sequences/Samples	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	45	41	51	61	15	29
Primer R GATCCTGTGAGACTCCTCGC	41	35	52	46	29	24
wild-type Subcortical areas at 1W						
Sequences/Samples	Subcor_1W_F	Subcor_1W_R	Subcor_1W_F	Subcor_1W_R	Subcor_1W_F	Subcor_1W_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	57	41	38	24	35	29
Primer R GATCCTGTGAGACTCCTCGC	31	32	32	32	30	28
wild-type Subcortical areas at 1M						
Sequences/Samples	Subcor_1M_F	Subcor_1M_R	Subcor_1M_F	Subcor_1M_R	Subcor_1M_F	Subcor_1M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	33	22	24	20	37	38
Primer R GATCCTGTGAGACTCCTCGC	17	27	23	23	34	37
wild-type Subcortical areas at 3M						
Sequences/Samples	Subcor_3M_F	Subcor_3M_R	Subcor_3M_F	Subcor_3M_R	Subcor_3M_F	Subcor_3M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	47	34	48	40	33	29
Primer R GATCCTGTGAGACTCCTCGC	37	28	36	26	31	31
wild-type Subcortical areas at 6M						
Sequences/Samples	Subcor_6M_F	Subcor_6M_R	Subcor_6M_F	Subcor_6M_R	Subcor_6M_F	Subcor_6M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	on
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	38	44	26	23	50	37
Primer R GATCCTGTGAGACTCCTCGC	34	43	30	13	25	37

In the case of primers specific to Grb10, they had existed in both WT and mutant Grb10KO$^{+/p}$ mice at all the developmental stages in both the subcortical areas (Table 3.1 and Table 3.3) and the cortex (Table 3.2 and Table 3.4). However, in the subcortical areas, the total number has been greater than in the cortex. In the subcortical areas, total numbers of primers specific to Grb10 in WT (Table 3.1) mice had shown double than mutant Grb10KO$^{+/p}$ mice (Table 3.3).

Table 3.2. Primers specific to β-geo insertion and Grb10 in the cortex in wild-type mice. Reports for detecting primers specific for inserted β-geo gene trap cassette and the Grb10 in the cortex (Cor) at all the 5 developmental stages (=3 at each stage) on the set of paired-end reads (F= forward reads, R= reverse reads) in wild-type mice using bash command line.

wild-type Cortex at E18.5						
Sequences/Samples	Cor_E18.5_F	Cor_E18.5_R	Cor_E18.5_F	Cor_E18.5_R	Cor_E18.5_F	Cor_E18.5_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	22	24	33	29	26	10
Primer R GATCCTGTGAGACTCCTCGC	21	14	16	15	13	19
wild-type Cortex at 1W						
Sequences/Samples	Cor_1W_F	Cor_1W_R	Cor_1W_F	Cor_1W_R	Cor_1W_F	Cor_1W_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	15	16	18	23	20	14
Primer R GATCCTGTGAGACTCCTCGC	18	12	17	17	20	17
wild-type Cortex at 1M						
Sequences/Samples	Cor_1M_F	Cor_1M_R	Cor_1M_F	Cor_1M_R	Cor_1M_F	Cor_1M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	20	11	18	12	8	6
Primer R GATCCTGTGAGACTCCTCGC	9	10	21	9	5	9
wild-type Cortex at 3M						
Sequences/Samples	Cor_3M_F	Cor_3M_R	Cor_3M_F	Cor_3M_R	Cor_3M_F	Cor_3M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	20	19	19	15	24	13
Primer R GATCCTGTGAGACTCCTCGC	11	18	10	13	14	18
wild-type Cortex at 6M						
Sequences/Samples	Cor_6M_F	Cor_6M_R	Cor_6M_F	Cor_6M_R	Cor_6M_F	Cor_6M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	0	0	0	0	0	0
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	20	11	13	11	18	20
Primer R GATCCTGTGAGACTCCTCGC	8	15	17	12	16	16

Table 3.3. Primers specific to β-geo insertion and Grb10 in the subcortical areas in mutant mice. Reports for detecting primers specific for inserted β-geo gene trap cassette and the Grb10 in the subcortical areas (Subcor) at all the 5 developmental stages (=3 at each stage) on the set of paired-end reads (F= forward reads, R= reverse reads) in mutant mice using bash command line.

Mutant Subcortical areas at E18.5						
Sequences/Samples	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_	Subcor_E18.5_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	31	31	30	20	1	3
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	20	28	17	19	27	25
Primer R GATCCTGTGAGACTCCTCGC	26	21	26	15	16	20
Mutant Subcortical areas at 1W						
Sequences/Samples	Subcor_1W_F	Subcor_1W_R	Subcor_1W_F	Subcor_1W_R	Subcor_1W_F	Subcor_1W_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	25	19	48	37	47	44
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	4	8	18	22	13	20
Primer R GATCCTGTGAGACTCCTCGC	10	6	17	14	16	12
Mutant Subcortical areas at 1M						
Sequences/Samples	Subcor_1M_F	Subcor_1M_R	Subcor_1M_F	Subcor_1M_R	Subcor_1M_F	Subcor_1M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	16	15	42	34	2	1
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	10	6	22	14	28	33
Primer R GATCCTGTGAGACTCCTCGC	10	8	16	13	33	20
Mutant Subcortical areas at 3M						
Sequences/Samples	Subcor_3M_F	Subcor_3M_R	Subcor_3M_F	Subcor_3M_R	Subcor_3M_F	Subcor_3M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	23	18	28	23	17	22
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	5	10	9	5	7	11
Primer R GATCCTGTGAGACTCCTCGC	7	4	11	4	7	7
Mutant Subcortical areas at 6M						
Sequences/Samples	Subcor_6M_F	Subcor_6M_R	Subcor_6M_F	Subcor_6M_R	Subcor_6M_F	Subcor_6M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	31	32	28	24	57	41
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	19	11	16	10	17	20
Primer R GATCCTGTGAGACTCCTCGC	9	6	11	15	13	15

Table 3.4. Primers specific to β-geo insertion and Grb10 in the cortex in mutant mice. Reports for detecting primers specific for inserted β-geo gene trap cassette and the Grb10 in the cortex (Cor) at all the 5 developmental stages (=3 at each stage) on the set of paired-end reads (F= forward reads, R= reverse reads) in mutant mice using bash command line.

Mutant Cortex at E18.5						
Sequences/Samples	Cor_E18.5_F	Cor_E18.5_R	Cor_E18.5_F	Cor_E18.5_R	Cor_E18.5_F	Cor_E18.5_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	5	9	4	9	32	29
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	25	14	22	23	24	21
Primer R GATCCTGTGAGACTCCTCGC	19	22	35	23	27	16
Mutant Cortex at 1W						
Sequences/Samples	Cor_1W_F	Cor_1W_R	Cor_1W_F	Cor_1W_R	Cor_1W_F	Cor_1W_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	6	8	10	8	13	17
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	3	3	12	9	15	14
Primer R GATCCTGTGAGACTCCTCGC	4	1	11	5	13	7
Mutant Cortex at 1M						
Sequences/Samples	Cor_1M_F	Cor_1M_R	Cor_1M_F	Cor_1M_R	Cor_1M_F	Cor_1M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	18	11	18	13	1	3
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	9	7	6	5	14	14
Primer R GATCCTGTGAGACTCCTCGC	2	2	6	6	16	4
Mutant Cortex at 3M						
Sequences/Samples	Cor_3M_F	Cor_3M_R	Cor_3M_F	Cor_3M_R	Cor_3M_F	Cor_3M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	11	10	17	10	30	16
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	9	10	4	13	13	4
Primer R GATCCTGTGAGACTCCTCGC	5	4	6	9	13	4
Mutant Cortex at 6M						
Sequences/Samples	Cor_6M_F	Cor_6M_R	Cor_6M_F	Cor_6M_R	Cor_6M_F	Cor_6M_R
bgeo_F2 CCGAGCGAAAACGGTCTGCG	24	19	11	10	22	24
bgeo_R2 CTTCCGCTTAGTGACAACG	0	0	0	0	0	0
Primer F CTGACCTGGAAGAAAGCAGC	4	6	13	14	12	16
Primer R GATCCTGTGAGACTCCTCGC	6	7	2	8	5	4

High quality RNA-Seq data allowed transcriptome profiling of Grb10KO$^{+/p}$ mice

In order to analyse the transcriptome data from 60 samples of midbrain and neocortex, the following pipeline was used (Figure 3.1). The FastQC application of the Galaxy version 0.67 was used to overview the range of quality values across all bases of the samples. To simplify overview of the FastQC analysis two samples were chosen

sample number one represents the subcortical areas of E18.5 mutant Grb10KO$^{+/p}$ (forward read), and sample number 55 represents the cortex of 3M old WT mouse (reverse read). The report shows that all the paired-end reads samples have a sequence length of 125 bp with zero poor quality sequences. An example of quality control describing the basic statistics of the raw RNA-Seq dataset in the subcortical areas (Table 3.5) and the cortex (Table 3.6) has been provided.

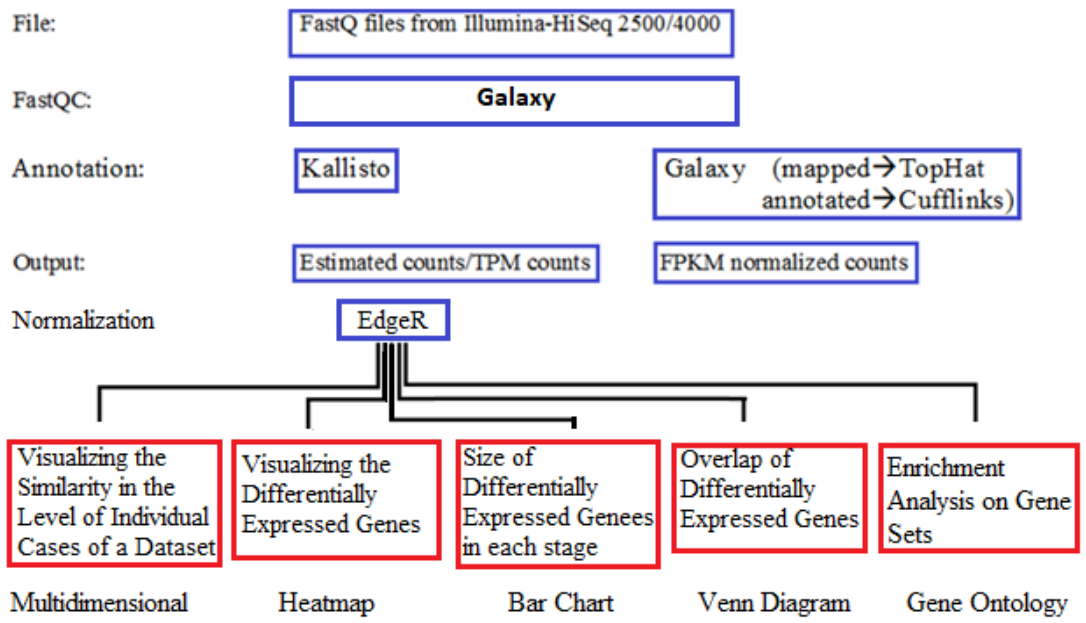


Figure 3.1. Bioinformatics analyses pipeline. A flowchart summarising the basic bioinformatic analysis of reads obtained from RNA-Seq datasets (WT & mutant Grb10KO$^{p/+}$), including the software components.

The quality scores in Phred scale for all samples had been above 23 (more than 99% base accuracy) with the majority (90%) above 25 as shown for the subcortical areas (Figure 3.2) and the cortex (Figure 3.3). All the sequencing adapters had been trimmed using the trimmomatic tool by the BGI Tech Solutions (Hong Kong) Co., Ltd as shown in Figure 3.4 for the subcortical areas and Figure 3.5 for the cortex.

Table 3.5. FastQC Basic Statistics. Quality control describing the basic statistics of the raw RNA-Seq dataset in one of the subcortical areas samples (Sample 1, E18.5 mutant Grb10KO$^{+/p}$ mouse, forward strand sequence).

Measure	Value
Filename	KO_E18.5_Subcor_S1_F. fq
Total sequence	23031118
Sequence flagged as poor quality	0
Sequence length	125
%GC	49

Table 3.6. FastQC Basic Statistics. Quality control describing the basic statistics of the raw RNA-Seq dataset in one of the cortex samples (Sample no. 55, 3M old WT mouse, reverse strand sequence).

Measure	Value
Filename	WT_3M_Cor_S55_R. fq
Total sequence	26805891
Sequence flagged as poor quality	0
Sequence length	125
%GC	49

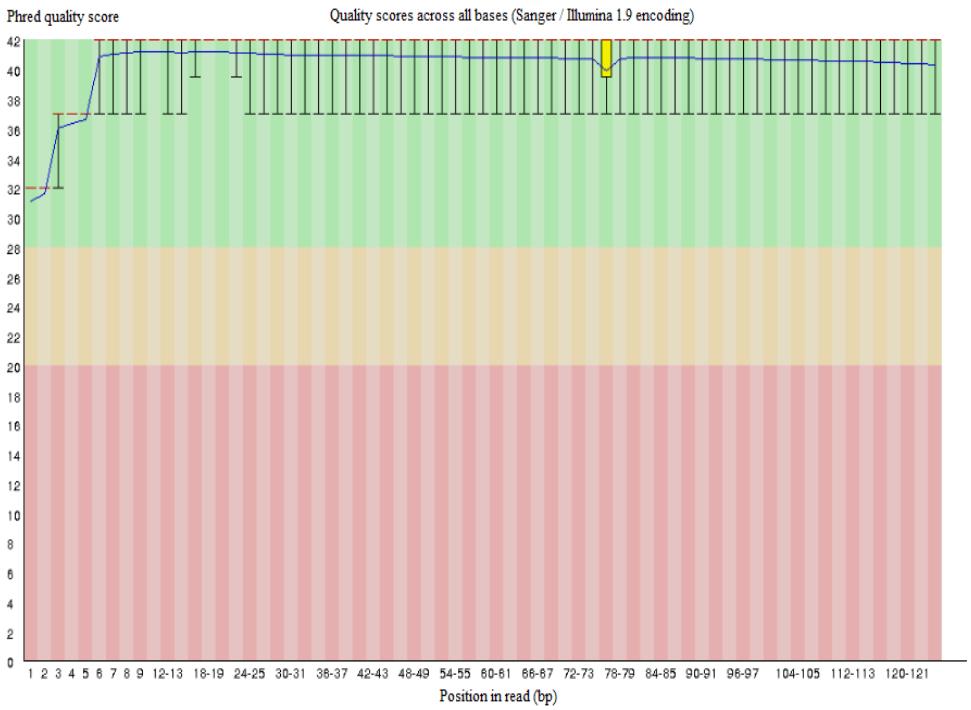


Figure 3.2. FastQC Per Base Sequence Quality. Quality control describing the sequence quality per base of the raw RNA-Seq dataset in one of the subcortical areas samples (Sample no. 1). The blue line represents the mean quality.

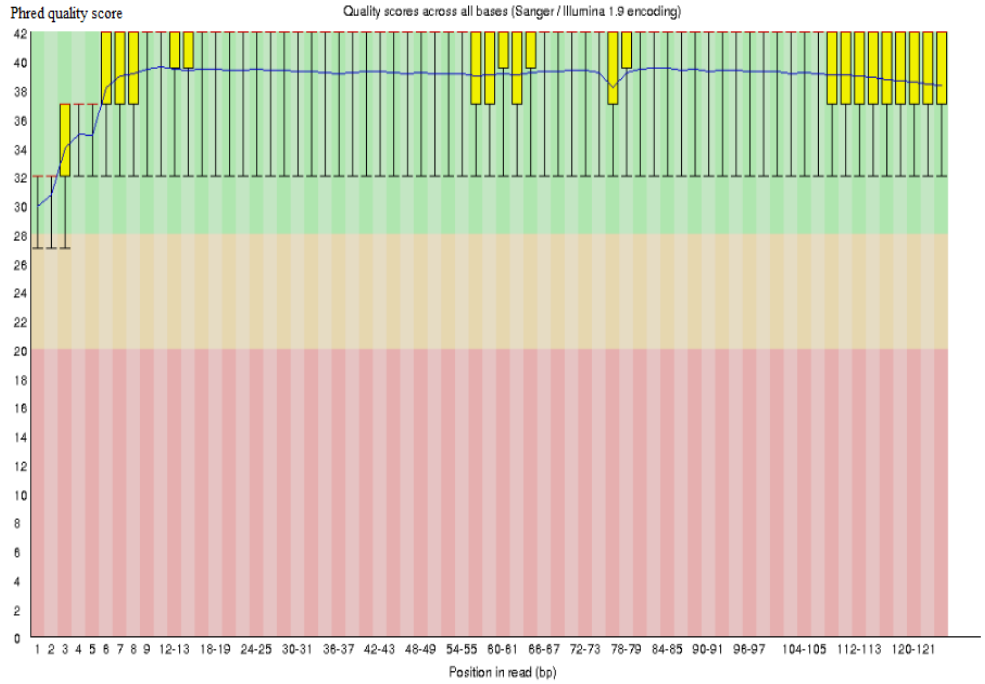


Figure 3.3. FastQC Per Base Sequence Quality. Quality control describing the sequence quality per base of the raw RNA-Seq dataset in one of the cortex samples (Sample no. 55). The blue line represents the mean quality.

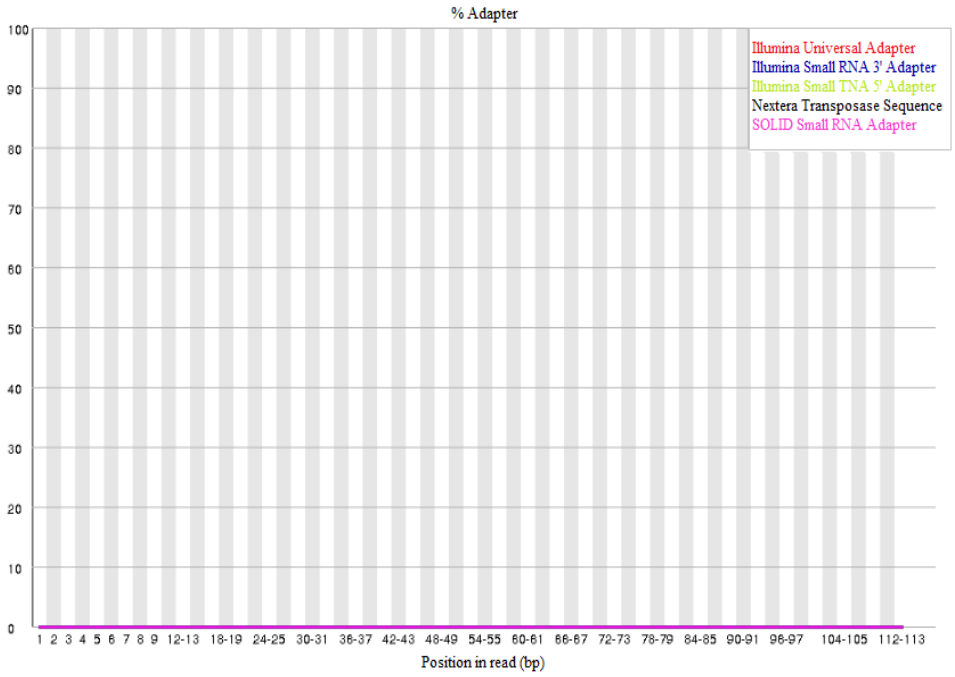


Figure 3.4. FastQC Adapter Content. Quality Control describing the adapter content in the raw RNA-Seq dataset in one of the subcortical areas samples (Sample no. 1).

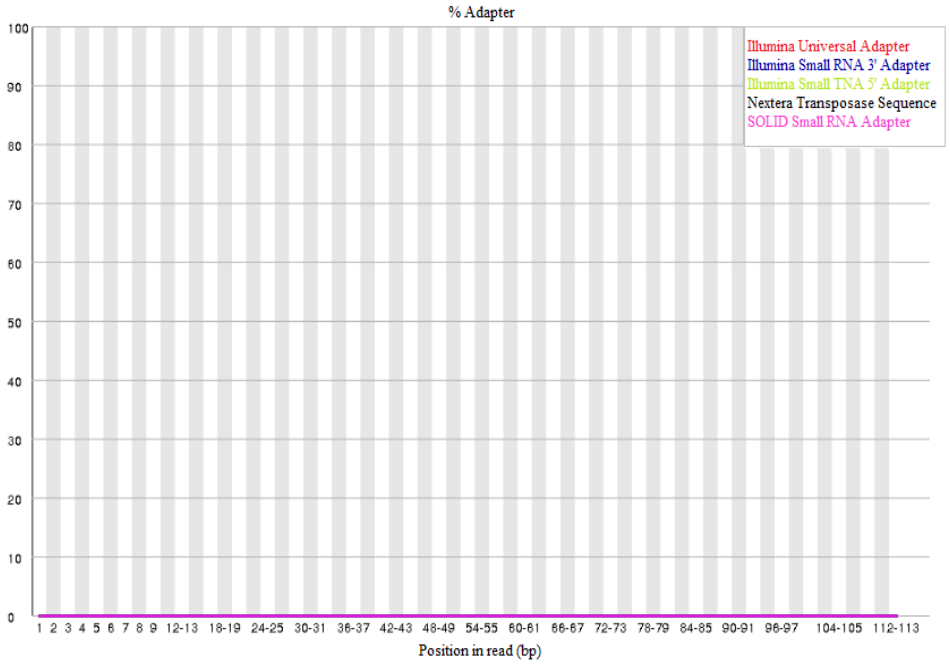


Figure 3.5. FastQC Adapter Content. Quality control describing the adapter content in the raw RNA-Seq dataset in one of the cortex samples (Sample no. 55).

Annotation of the transcripts was performed in Linux operating system using Kallisto program. Output file of abundances.tsv was used for further analysis. A total number of 103215 transcriptomes and 33366 genes had been obtained for each brain tissue sample (Table 3.7). Bash command line had been used to detect the total number of reads ranged from 40251638 to 66811350 (Table 3.7).

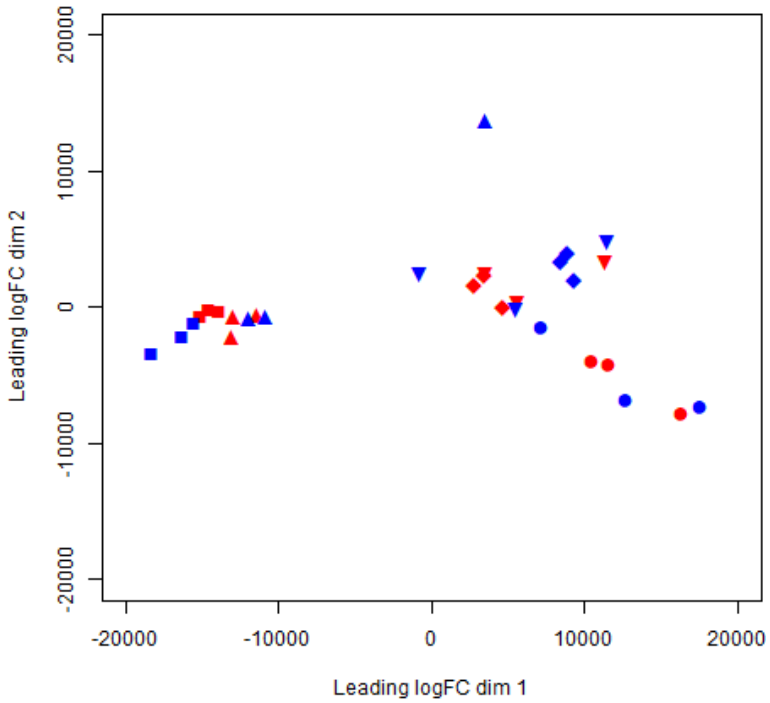
Overall similarity in Grb10KO and wild-type transcriptome profiles

MDS is a method used for treating data with various characteristics scattered in high-dimensional areas (Lopes, Machado and Galhano, 2017). For this project, it is used to explore the expression similarity between the samples in the subcortical areas and the cortex. MDS revealed 2 clusters in the subcortical areas (Figure 3.6A) and 2 clusters in the cortex (Figure 3.6B) in which the samples were divided per age group between WT and mutant Grb10KO$^{+/p}$ (red colour represents WT type while blue colour represents mutant Grb10KO$^{+/p}$). The cortex seems to show no differences in expression level in both WT and mutant Grb10KO$^{+/p}$ at any age stages, wherein subcortical areas there is slight difference expression between the WT and mutant Grb10KO$^{+/p}$ at adult stage mainly at 3 months old.

Table 3.7. Annotation. A total number of annotated genes and transcripts using Kallisto and the total number of reads (paired-end reads) on each developmental stage (=3) in the subcortical areas and the cortex.

Total number of annotated transcripts =103215 **Total number of annotated genes =**33366		
Total number of reads (forward and reverse reads) in each replica (=3) on each developmental stage:		
E18.5 (Subcortical)	n1 =55488680 n2 =58536030 n3 =40919044	n1 =46062236 n2 =42836388 n3 =48698068
1W old (Subcortical)	n1 =55070770 n2 =48002886 n3 =41649884	n1 =54732926 n2 =48870784 n3 =47344300
1M old mice (Subcortical)	n1 =41841970 n2 =50157926 n3 =66811350	n1 =42093222 n2 =51438074 n3 =55839318
3M old mice (Subcortical)	n1 =48687052 n2 =52360956 n3 =51535504	n1 =44875358 n2 =41465716 n3 =42212470
6M old mice (Subcortical)	n1 =50105648 n2 =40251638 n3 =52366034	n1 =57253950 n2 =57422230 n3 =47074290
E18.5 (Cortex)	n1 =52805624 n2 =51089404 n3 =52672758	n1 =56587962 n2 =54097536 n3 =50536320
1W old mice (Cortex)	n1 =40865120 n2 =46225890 n3 =54634408	n1 =51604590 n2 =41116926 n3 =57511948
1M old mice (Cortex)	n1 =55722972 n2 =54399812 n3 =53078102	n1 =51400744 n2 =47038912 n3 =49394700
3M old mice (Cortex)	n1 =53611782 n2 =55760778 n3 =51471176	n1 =47997350 n2 =52266490 n3 =48686128
6M old mice (Cortex)	n1 =53882884 n2 =52606502 n3 =52773798	n1 =47156654 n2 =56582966 n3 =58829282

A. Subcortical areas



B. Neocortex

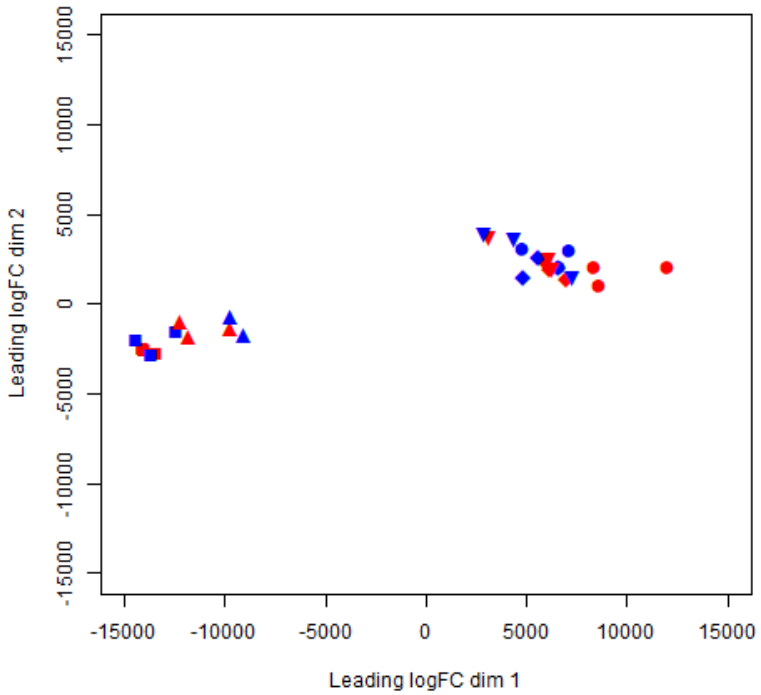


Figure 3.6. Multidimensional scaling plot of the normalised data samples of Grb10KO$^{p/+}$ and wild-type mice. Red colour represents WT mice while blue colour represents mutant mice. Different developmental age group represented by certain symbols: E18.5 ■ ■, 1W ▲ ▲, 1M ● ●, 3M ◆ ◆ and 6M ▼ ▼. A) Samples from subcortical areas and B) neocortex.

Differentially expressed genes found in subcortical areas and neocortex in Grb10KO

The differential expression analysis was carried out using the edgeR package accessible in the *R* software (Robinson et al., 2010). The analysis was performed on each developmental stage independently and the number of down and up-regulated differentially expressed genes obtained for each stage (FDR < 0.05) was quantified (Figure 3.7). The overall observation on the amount of differentially expressed genes of both up- and down-regulated through developmental stages showed the higher expression of differentially expressed genes in the subcortical areas (139 genes) and the cortex (26 genes) is at 3M old (Figure 3.7). At 6M the differentially expressed genes of both up- and down-regulated decrease (Figure 3.7) suggesting a fluctuation in the expression level at the adult stage. This result suggests that most transcriptional alterations take place at the stage of 3M old. Differentially expressed genes at adult stages (3M and 6M) old mice in both tissues were then clustered with a conventional heat map analysis Figure 3.8. As expected, Grb10 is down-regulated in several comparisons.

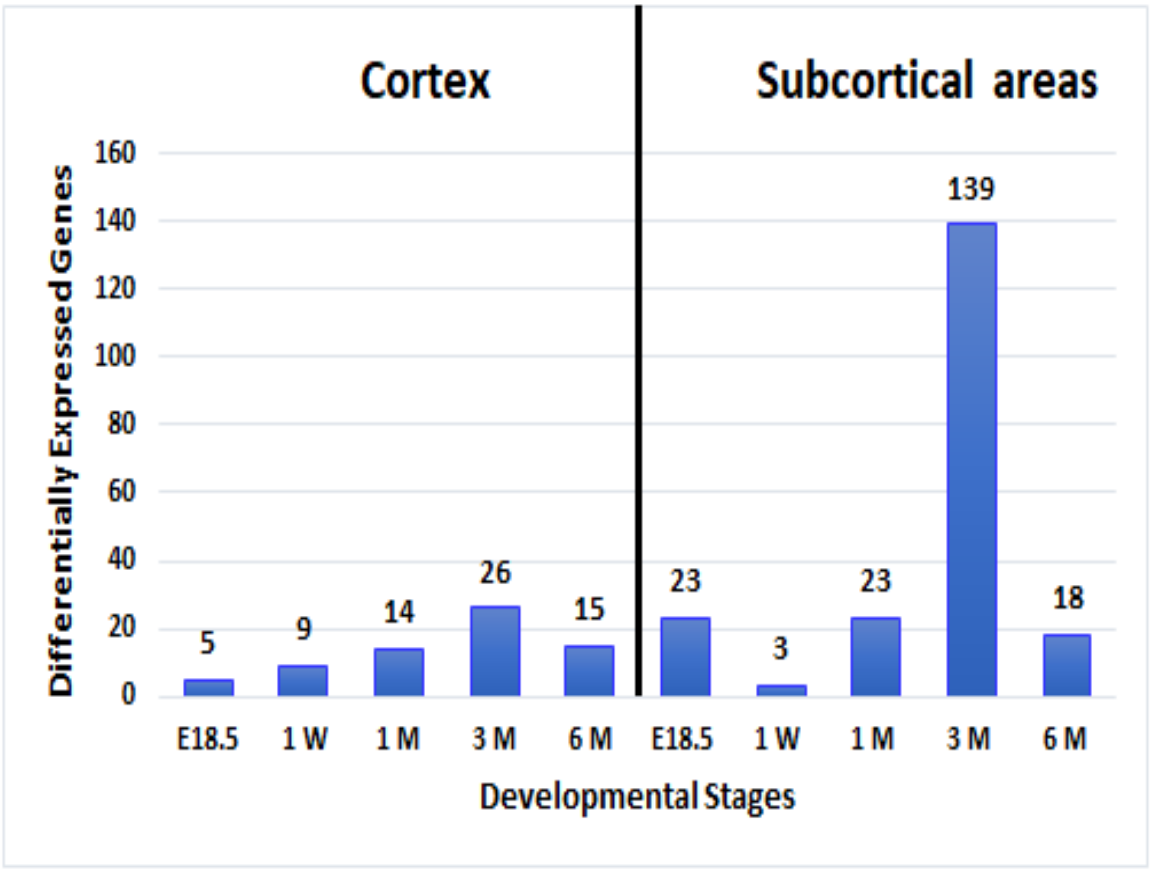
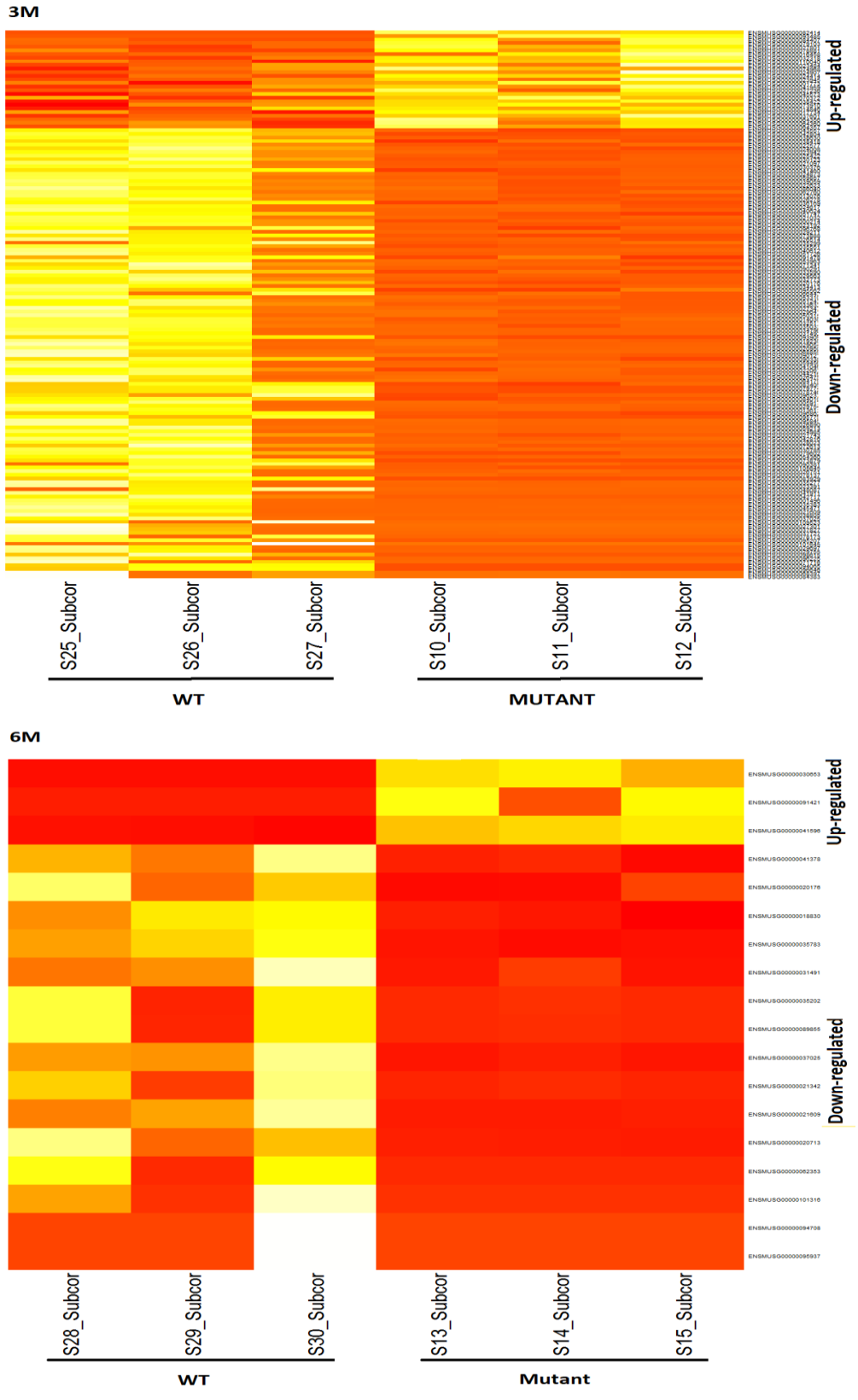


Figure 3.7. Bar chart for the number of differentially expressed genes/developmental stages/tissue. The bar chart shows the actual number of up- and down-regulated differentially expressed genes for each developmental stage represented in the subcortical areas and the cortex.

Heatmap of genes clusters in the subcortical areas (A) and the cortex (B).

A. Subcortical areas



B. Cortex

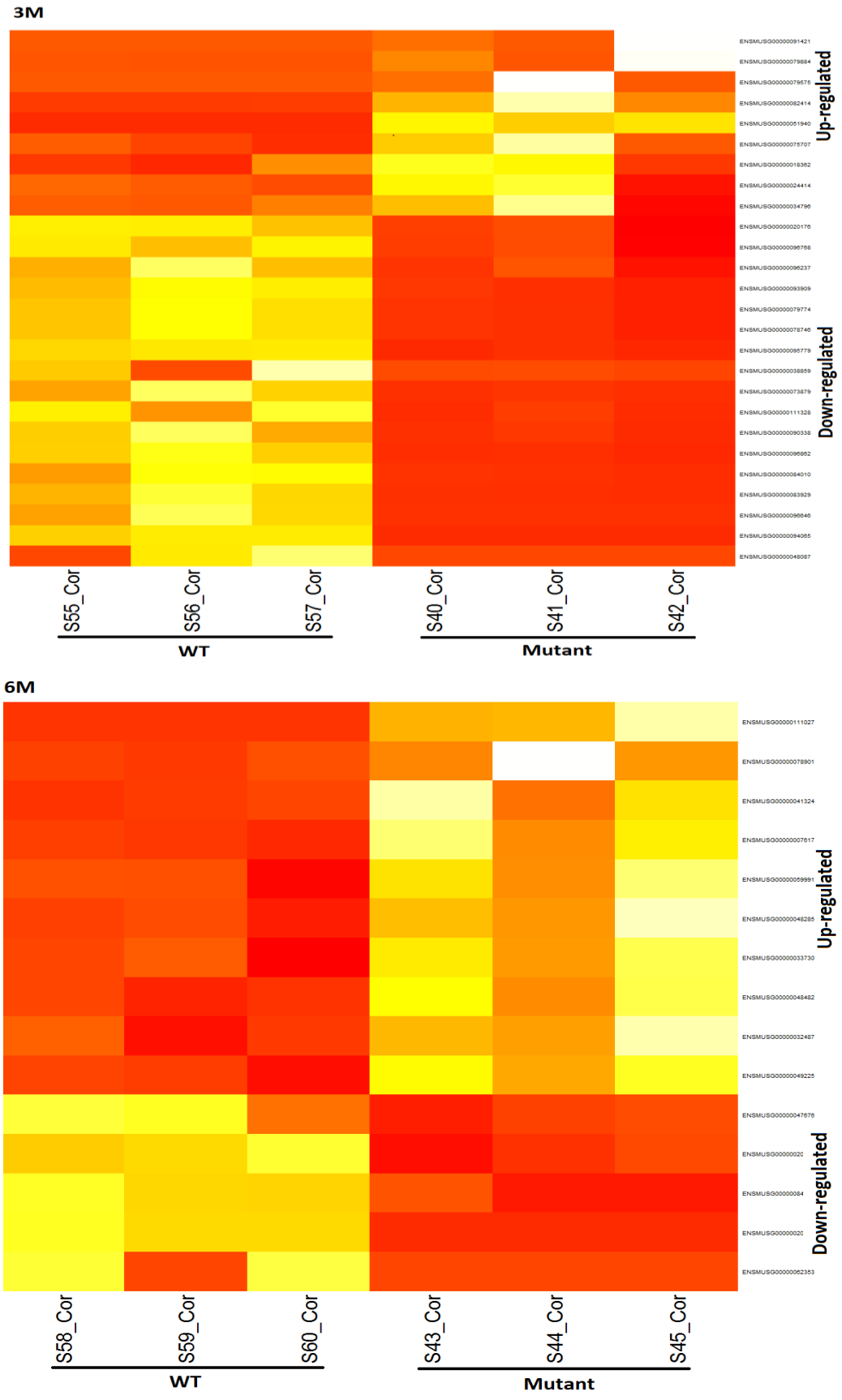

3M
Up-regulated
Down-regulated
S55_Cor
S56_Cor
S57_Cor
S40_Cor
S41_Cor
S42_Cor
WT
Mutant
6M
Up-regulated
Down-regulated
S58_Cor
S59_Cor
S60_Cor
S43_Cor
S44_Cor
S45_Cor
WT
Mutant

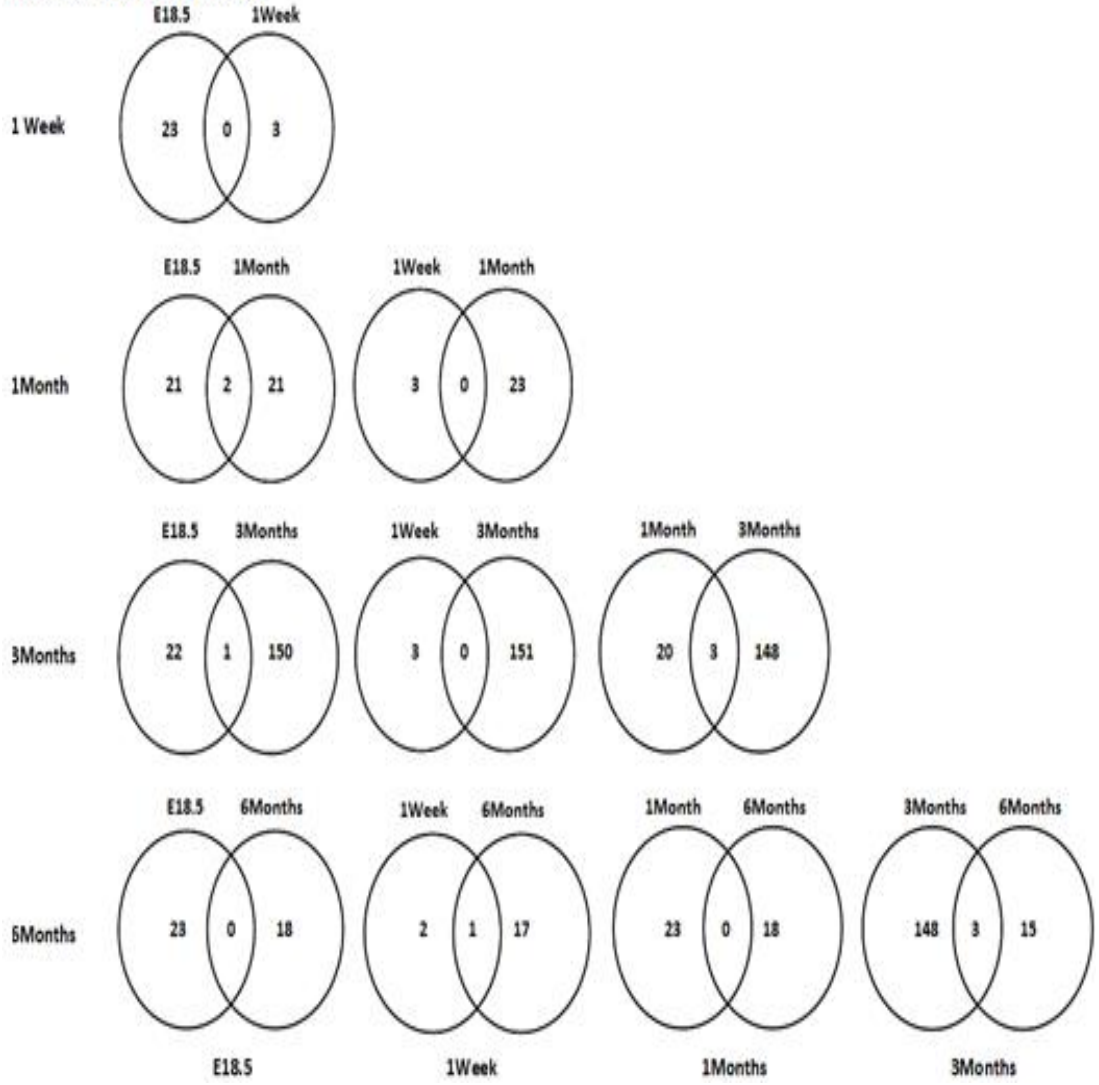
Figure 3.8. Heatmap of genes clusters in the subcortical areas (A) and the cortex (B) of the WT and mutant Grab10KO$^{+/p}$ at 3 and 6 months old mice. Heatmap of genes clusters as expressed in the WT and mutant Grab10 at 3 and 6 months old mice in the subcortical areas (A) and in the cortex (B. The column in each diagram represents the total of 6 replica samples (labelled bottom): 3 replicas of WT (left) and 3 mutant Grb10KO$^{+/p}$ (right). The cluster of Ensembl gene ID's is labelled on the right. The colour bar indicates the colour coding of gene expression for the Figure (PDR < 0.05), it ranged from the red colour that indicates gene expression peaks to the white colour that indicates lower or null gene expression. The top cluster genes are up-regulated (log FC > 0); the bottom cluster genes are down-regulated (log FC < 0). Grb10 at both developmental stages in both tissue samples is down-regulated.

To check the consistent core of differentially expressed genes throughout adult development, overlap measurement of differentially expressed genes across the different developmental stages was made (Figure 3.9) to identify those genes that are either consistently expressed in most of the groups or are specific to a single developmental stage. The pairwise comparisons between stages revealed that there is null or no considerable amount of differentially expressed genes shared across the developmental groups in both the subcortical areas (Figure 3.9A and Table 3.8) and the cortex (Figure 3.9B and Table 3.8) except 1-3 genes.

We examined the overlapping differentially expressed genes between subcortical areas and the cortex we found that both tissues showed changes in opposite direction: up-regulated in one stage and down-regulated in other except in the subcortical areas as the Gm28635 (predicted gene 28635) was up-regulated at E18.5 and 1M while Gm12663 (Predicted gene 12663) was down-regulated at 1W and 6M as well as the Slc6a3, Foxa2 and Grb10 were down-regulated at 3M and 6M. In the cortex, the Gm4202 (Predicted gene 4202) was up-regulated at 1M and 3M while Grb10 was down-regulated at 3M and 6M (Table 3.8). We also found that only two genes were differentially expressed in both tissues which are Rps2 between E18.5 and 1M and Grb10 between 3M and 6M old mice (Table 3.8).

GO enrichment for the overlapped genes was also made. It revealed no adjusted significant values. Furthermore, analysis of shared transcript changes between the subcortical areas and the cortex at all five-time points has been conducted which revealed 29 shared genes (Figure 3.10) in which most are predicted genes namely: Grb10, Gh, Mrpl27, Cpne7, Avp, Rps2, Hmga1, Fam205a4, Gdf1, CT868723.1, CR974586.6, Gm4737, Gm10052, Gm15772, Gm5859, Gm13298, Gm15920, Gm10600, Gm13302, Gm17081, Gm6158, Gm4202, Gm3883, Gm21541, Gm2163, Gm47283, Gm13301, Gm12663, Gm28635.

A. Subcortical areas



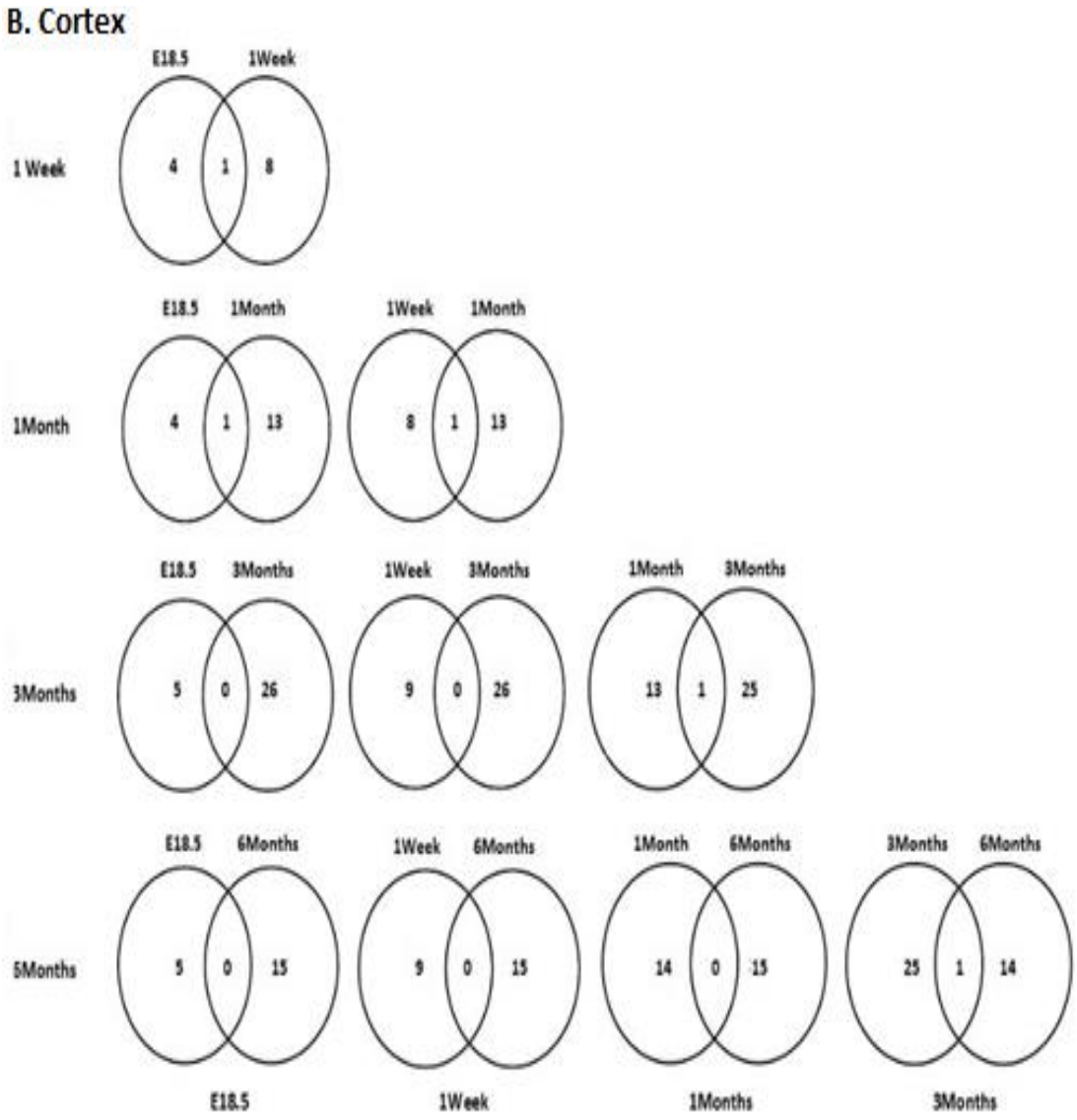


Figure 3.9. Venn diagram summarising the overlap of differentially expressed genes in the subcortical areas & the cortex. Pairwise comparison of differentially expressed genes in the subcortical areas (A) & the cortex (B) between the different developmental stages. The labels on the left and bottom of both Figures (A) & (B) indicate the different developmental stages of the mice samples. There is null or no considerable amount of differentially expressed genes shared across the developmental groups.

Table 3.8. Direction of pairwise differentially expressed genes between each developmental stage in both brain tissue. The direction of overlapped differentially expressed genes (down- or up-regulated) between the developmental stages in the subcortical areas and the cortex.

Developmental Stages of Overlapping	Gene Name	Down-regulated Developmental Stage	Up-regulated Developmental Stage
In the Subcortical areas			
E18.5 and 1M	1- Gm28635 (predicted gene 28635)	-	E18.5 and 1M
	2- Rps2 (Ribosomal protein S2**)**	1M	E18.5
E18.5 and 3M	Gm28635 (predicted gene 28635)	3M	E18.5
1W and 6M	Gm12663 (Predicted gene 12663)	1W and 6M	-
1M and 3M	1- Pmch (Pro-melanin-concentrating hormone)	3M	1M
	2- Hcrt (Orexin Orexin-A Orexin-B)	3M	1M
	3- Gm28635 (Predicted gene 28635)	3M	1M
3M and 6M	1- Grb10 (Growth factor receptor-bound protein 10)	3M and 6M	-
	2- Slc6a3 (Sodium-dependent dopamine transporter)	3M and 6M	-
	3- Foxa2 (Mus musculus forkhead box A2)	3M and 6M	-
In the Cortex			
E18.5 and 1W	Gm10052 (Predicted pseudogene 10052)	1W	E18.5
E18.5 and 1M	Rps2 (Ribosomal protein S2)	1M	E18.5
1M and 3M	Gm4202 (Predicted gene 4202)	-	1M and 3M
3M and 6M	Grb10 (Growth factor receptor-bound protein 10)	3M and 6M	-

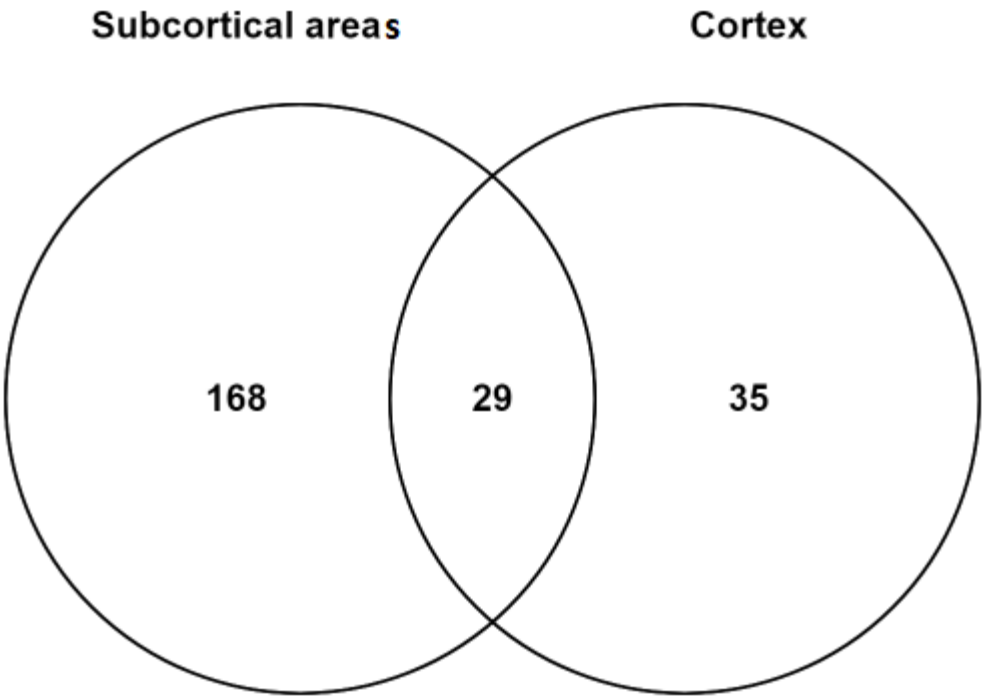


Figure 3.10. Venn diagram summarising the overlap of differentially expressed genes between the subcortical areas & the cortex. Pairwise comparison of differentially expressed genes between the subcortical areas & the cortex across all the 5 developmental stages.

Gene ontology enrichment analysis

The GO enrichment analysis showed an extensive list of gene ontology identifiers (GOIDs) for both sets biological process (=11204) and cellular component (=1500) which can be reduced by the gene count hence limiting the number of genes associated with GO categories. So, to avoid categories with only one or two genes associated comes up a significant causing difficulty for interpretation, the threshold of 10 genes related to the GO term was made, and the list of GOID was reduced for biological process n =2341 and for cellular component n =481.

Gene ontology enrichment analysis reveals that down-regulated differentially expressed genes in the subcortical areas targeted the biological processes mostly associated with the body response to drugs, hormones and other substances, development of the brain region and central nervous system regulation, cell development and regulation, development of the embryo and some vital organs & muscles, regulation of protein kinase and phosphorylation, ion transportation, regulation of transcription, they also associated with behaviours such as social

behaviour, locomotor, learning & memory, and finally with some signalling pathway such as Wnt, Bone morphogenetic proteins (BMP), Notch and chemokine (Figure 3.11A and supplementary Figure S3.1). While the significant GO categories of cellular component set related to neurons and transmission between them such as GABA-A receptor complex, synapse, postsynapse, dendrite, axon, its terminal bouton, plasma membrane, rough endoplasmic reticulum and extracellular region (Figure 3.12A) No significant categories found for down-regulated differentially expressed genes in the cortex in the both sets biological process and cellular component.

For up-regulated differentially expressed genes in the subcortical areas the biological processes associated with categories related to eating & feeding behaviour, response to ethanol hyperoxia & hypoxia, ATP synthesis & biosynthesis and neuropeptide signalling pathway (Figure 3.11B). While the significant GO categories of the cellular component set related to neurons such as perikaryon and neuronal cell body, cell junction, secretory granules in addition to categories related to the mitochondria such as mitochondrial proton-transporting ATP synthase, mitochondrial membrane and mitochondrion (Figure 3.12B).

While in the cortex significant categories of the biological processes for up-regulated differentially expressed genes associated with cell development, DNA transcription & gene expression, regulate cell proliferation, apoptotic process, vasoconstriction as well some signalling pathway such SMAD & MAPK cascade and response to cyclic compound, drug & ethanol (Figure 3.11C). The set of cellular component showed significant categories related to postsynapse, secretory granule, cytoplasmic membrane, ribosome and extracellular region (Figure 3.12C).

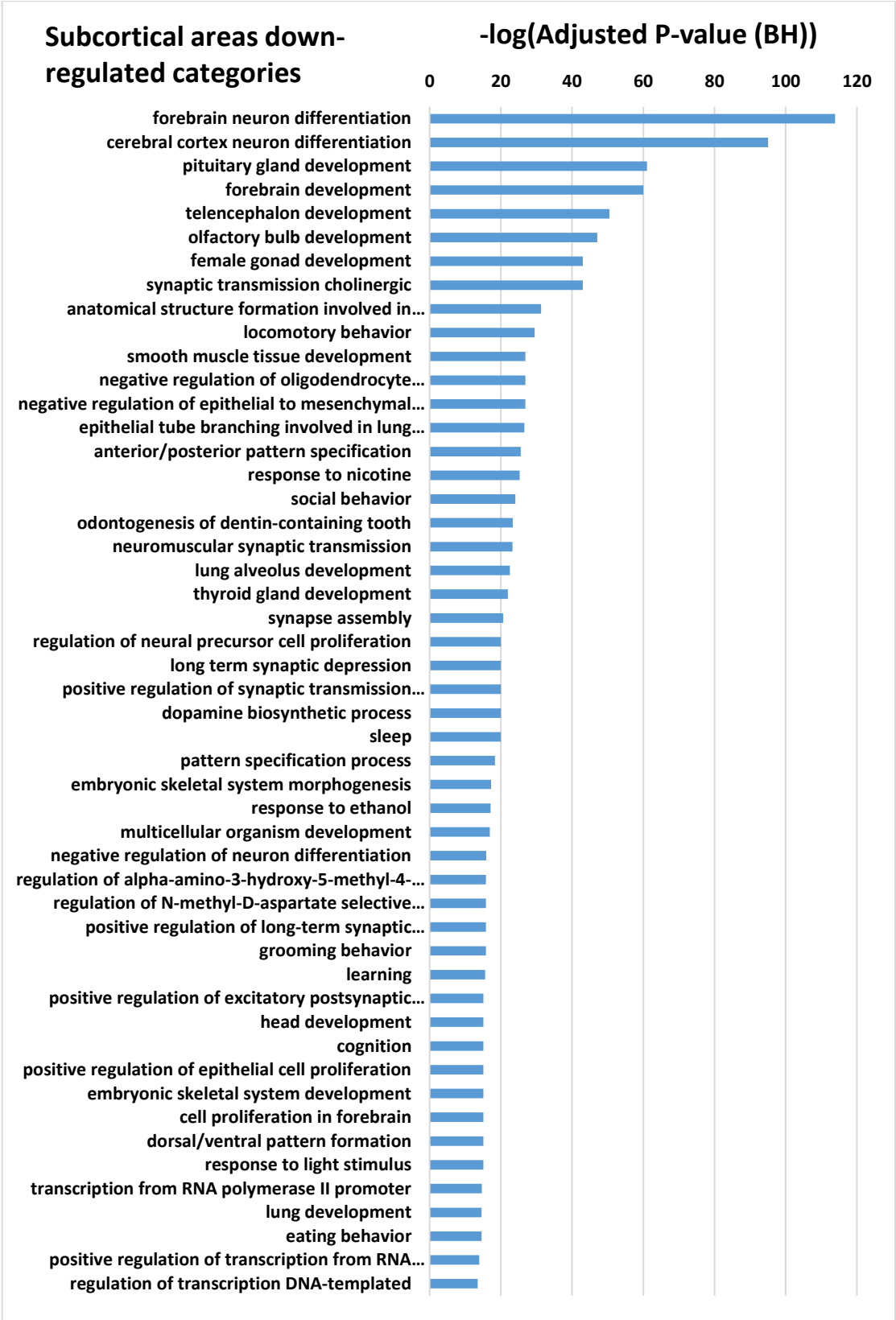
In this project since the highest number of gene expression appeared at 3 months old mice (Figure 3.7). Therefore, Gene ontology enrichment analysis was conducted at this stage to confirm and determine the region of the brain the dominant behaviour is conducted. At the stage of 3 months old, the result showed no identified significant GO enrichment categories in the cortex while the biological process for down-regulated differentially expressed genes in the subcortical areas confirmed the dominant behaviour that includes social behaviour, grooming behaviour and adult behaviour (Figure 3.13A, Table 3.9 and supplementary Figure S3.2). The biological process for up-regulated differentially expressed genes in the subcortical areas showed no behaviour categories they showed only biosynthesis and synthesis of

ATP, proton and ion transport, neuropeptide signalling pathway and nervous system development (Figure 3.13B).

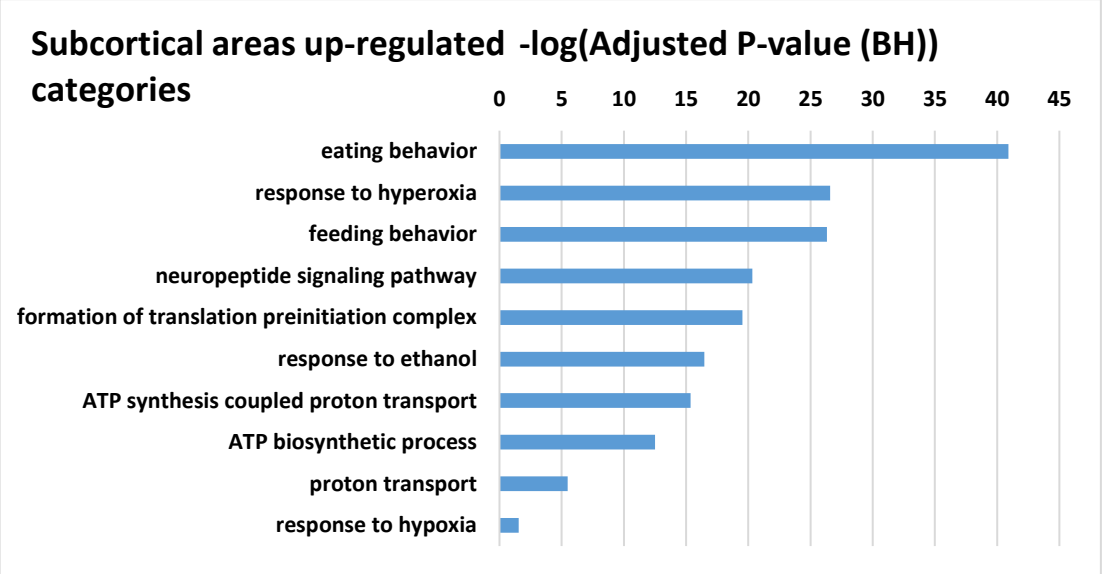
The cellular component for down-regulated differentially expressed genes in the subcortical areas showed categories related to neurons and transmission between them such as GABA-A receptor complex, synapse, postsynapse, dendrite, axon and terminal bouton (Figure 3.14). The cellular component for up-regulated differentially expressed genes in the subcortical areas also showed categories of neurons such as perikaryon, neuronal cell body and axon in addition to categories related to the mitochondria such as mitochondrial proton-transporting ATP synthase, mitochondrial membrane and mitochondrion (Figure 3.14). Gene ontology enrichment analysis revealed that each module targets a separate set of biological functions and cellular components, with rare or few functional overlap between modules.

There are 7 significant GO terms shared between the subcortical areas and the cortex with the biological process and 3 significant GO terms shared with the cellular component (Table 3.10 and supplementary Table S3.1).

A. GO categories of biological process for down-regulated differentially expressed genes in the subcortical areas. *Only top 50 GO categories are represented.



B. GO categories of biological process for up-regulated differentially expressed genes in the subcortical areas.



C. GO categories of biological process for up-regulated differentially expressed genes in the cortex.

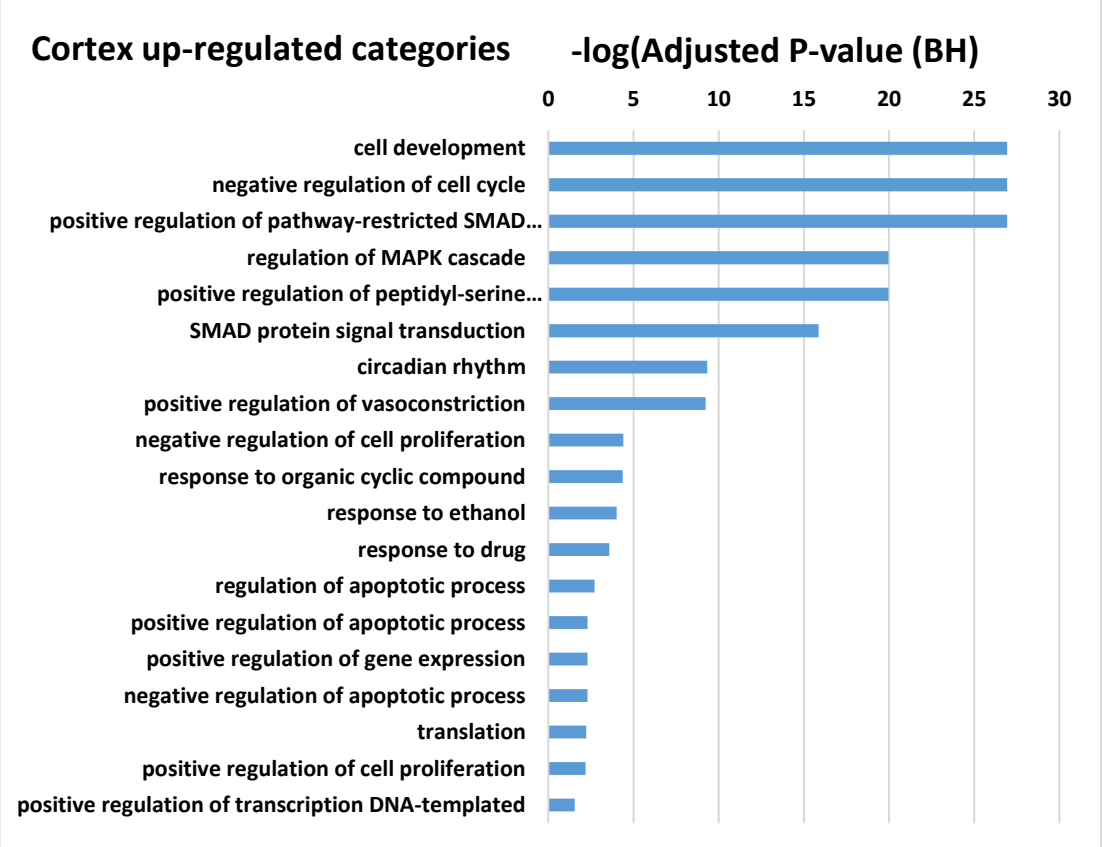
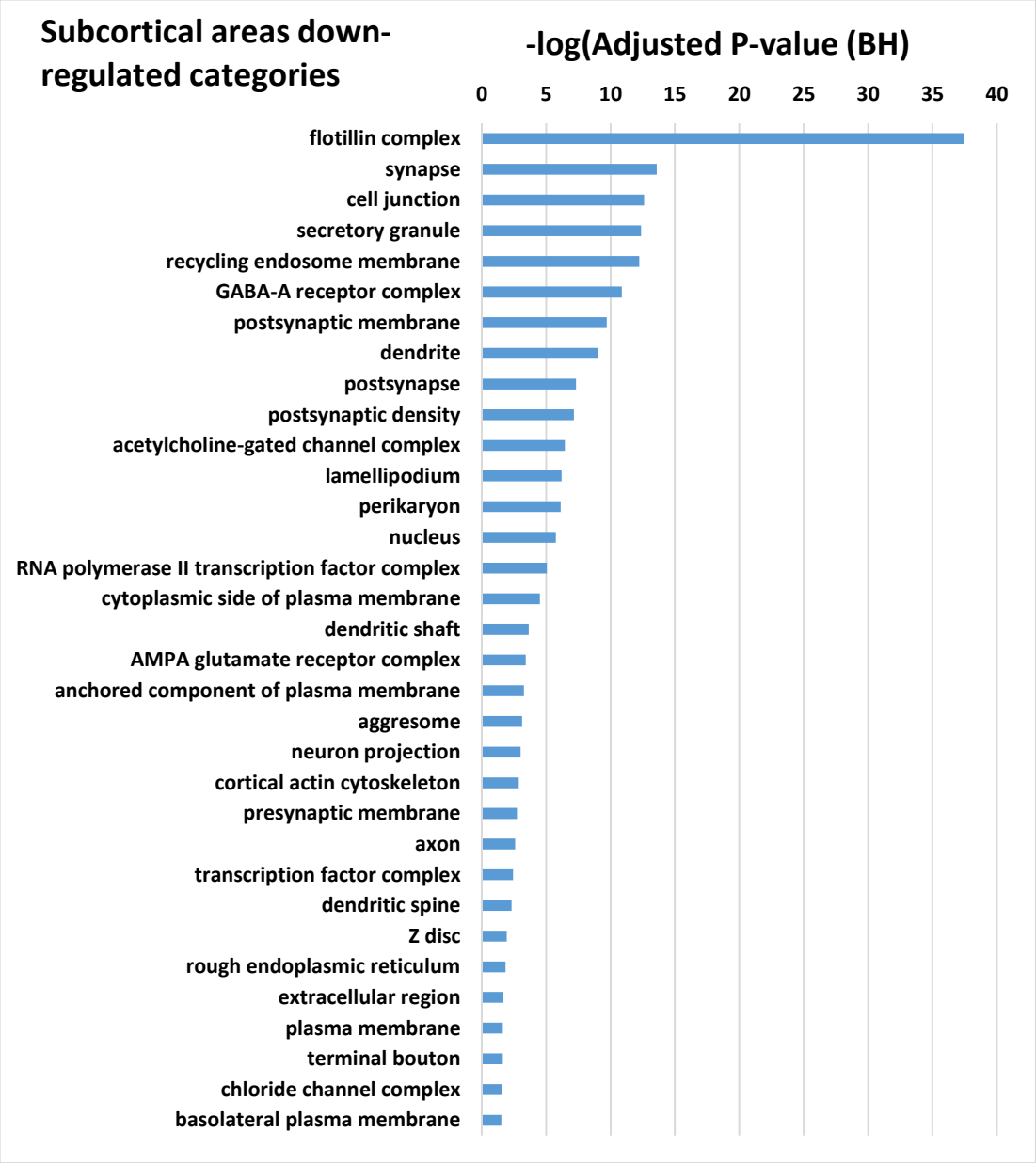


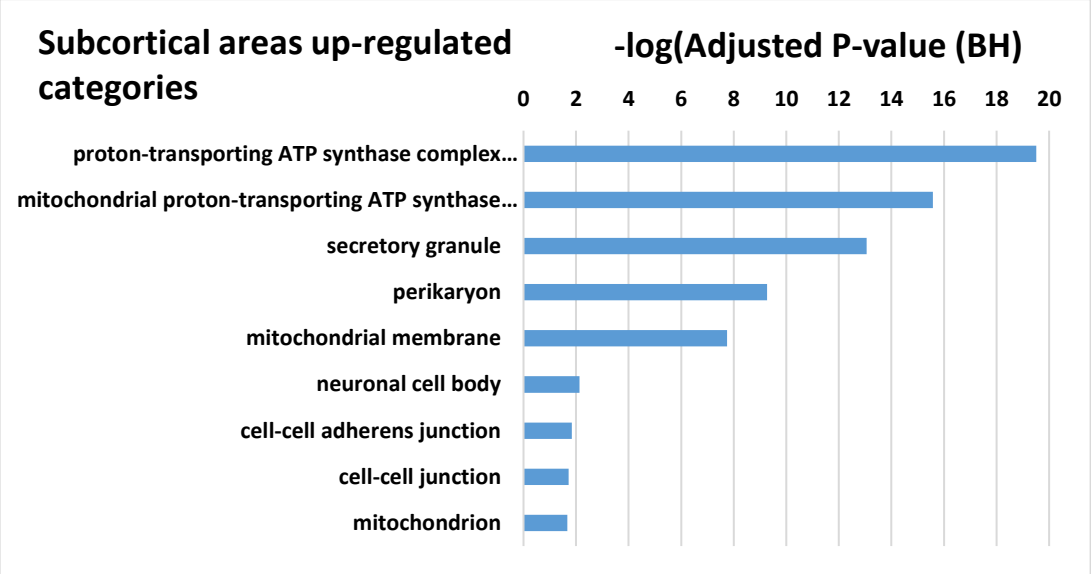
Figure 3.11. Bar plot showing the number of significantly enriched GO terms. Statistical significance in the enrichment of biological process GO terms was numerically assessed by

comparing with 10 000 equally sized random samples of genes for down- and up-regulated differentially expressed genes in the subcortical areas (A & B) and up-regulated differentially expressed genes in the cortex (C). The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

A. GO categories of cellular component for down-regulated differentially expressed genes in the subcortical areas.



B. GO categories of cellular component for up-regulated differentially expressed genes in the subcortical areas.



C. GO categories of cellular component for up-regulated differentially expressed genes in the cortex.

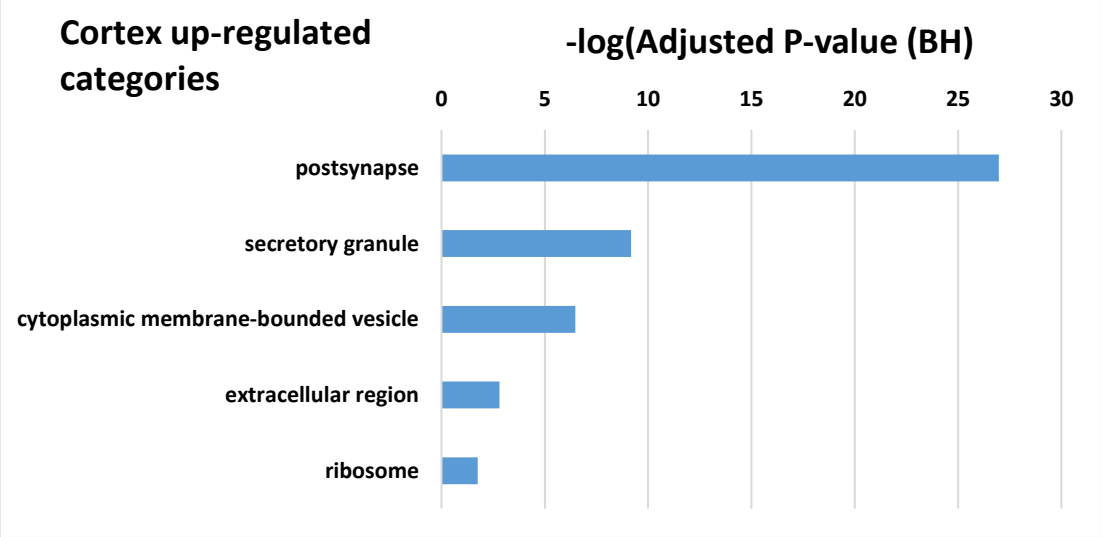


Figure 3.12. Bar plot showing the number of significantly enriched GO terms. Statistical significance in the enrichment of cellular component GO terms was numerically assessed by comparing with 10 000 equally sized random samples of genes for down- and up-regulated differentially expressed genes in the subcortical areas (A & B) and up-regulated differentially expressed genes in the cortex (C). The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

A. GO categories of biological process for down-regulated differentially expressed genes in the subcortical areas at 3 months old mice. *represented top 50 GO categories only.

B. GO categories of biological process for up-regulated differentially expressed genes in the subcortical areas at 3 months old mice.

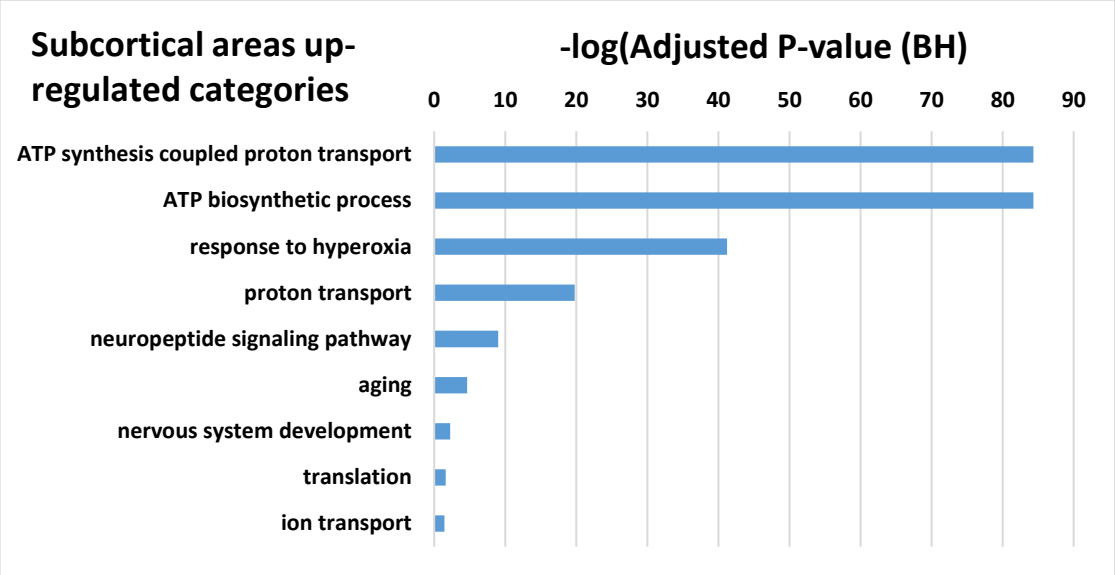
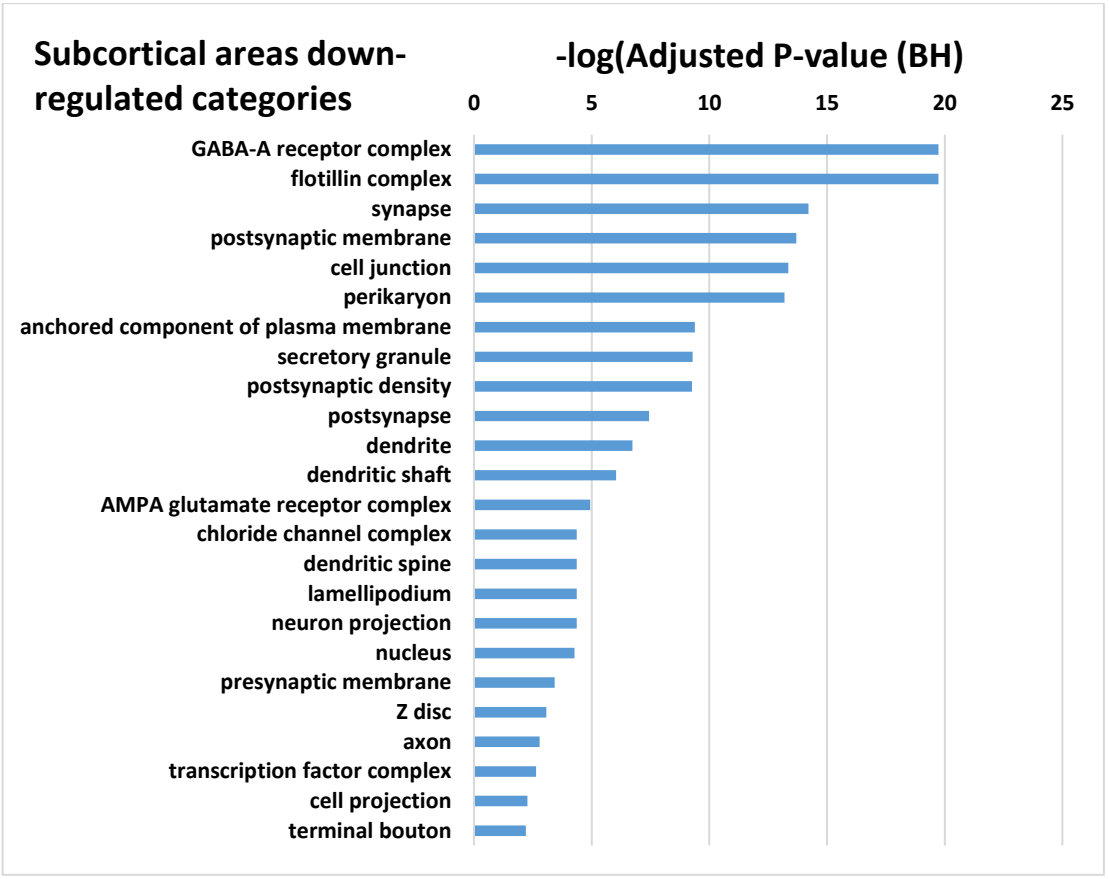


Figure 3.13. Bar plot showing the number of significantly enriched GO terms. Statistical significance in the enrichment of biological process GO terms was numerically assessed by comparing with 10 000 equally sized random samples of genes for down- and up-regulated differentially expressed genes in the subcortical areas at 3 months old mice (A & B). The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

Table 3.9. GO categories confirming the dominant behaviour in biological process for down-regulated differentially expressed genes in the subcortical areas at 3 months old mice.

GOID	GO Term	Adjusted P-value
GO:0035176	social behaviour	9.18E-19
GO:0030534	adult behaviour	3.91E-11
GO:0007625	grooming behaviour	5.49E-16
GO:0042711	maternal behaviour	6.55E-14
GO:0007626	locomotory behaviour	3.19E-38
GO:0042755	eating behaviour	3.11E-16
GO:0007631	feeding behaviour	3.91E-11
GO:0014823	response to activity	3.93E-08
GO:0007611	learning or memory	4.22E-07
GO:0007613	memory	1.02E-12
GO:0007612	learning	3.05E-11
GO:0030431	sleep	1.38E-20

A. GO categories of cellular component for down-regulated differentially expressed genes in the subcortical areas at 3 months old mice.



B. GO categories of cellular component for up-regulated differentially expressed genes in the subcortical areas at 3 months old mice.

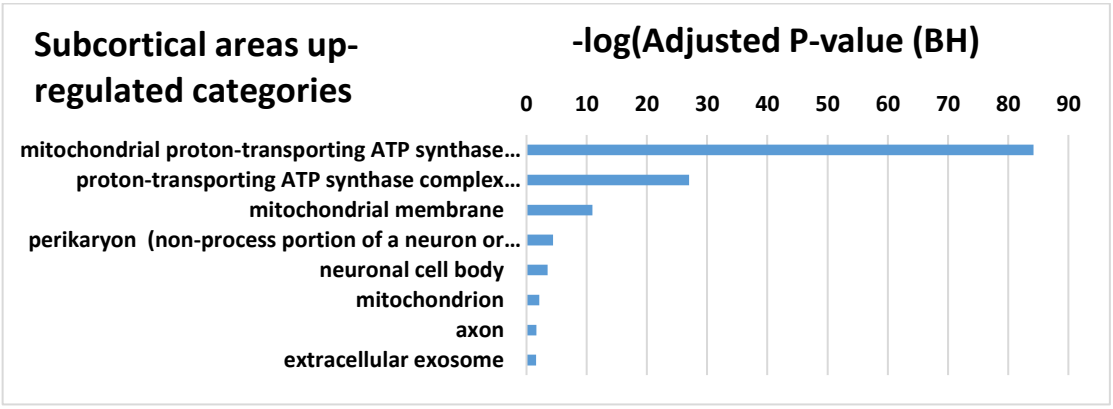


Figure 3.14. Bar plot showing the number of significantly enriched GO terms. Statistical significance in the enrichment of cellular component GO terms was numerically assessed by comparing with 10 000 equally sized random samples of genes for down- and up-regulated differentially expressed genes in the subcortical areas at 3 months old mice (A & B). The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

Table 3.10. Shared significant GO terms of biological process and cellular component between the subcortical areas and the cortex.

GO terms_Biological process	
GO:0008285	negative regulation of cell proliferation
GO:0010628	positive regulation of gene expression
GO:0014070	response to organic cyclic compound
GO:0042493	response to drug
GO:0045471	response to ethanol
GO:0045893	positive regulation of transcription DNA-templated
GO:0048468	cell development
GO terms_Celluar Component	
GO:0005576	Extracellular region
GO:0030141	Secretory granule
GO:0098794	postsynapse

Discussion

RNA-Seq data become a popular tool to address biological problems that can provide by using computational methods a complete picture of transcriptome facilitating the annotation and quantification of all genes in any sample. Since the mice in this project are originally the same generated mice reported by Garfield et al. (2011), the presence of the same primers specific to the β-geo insertion (probe A, Figure 2.1) assures that they are the same generated mice.

The presence of the primers specific to Grb10 in the cortex of the wild-type, although the total number is less than the subcortical areas, suggest that Grb10 also expressed in the cortex, but due to lower level of Grb10 expression in the cortex the *LacZ* neomycin2 gene-trap cassette which had inserted within Grb10 exon 8 (Figure 2.1) could not track it. In addition of that the presence of the primer specific to the β-geo insertion in both tissues of the mutant Grb10KO$^{+/p}$ in the cortex and the subcortical areas with a double number of this primer in the subcortical areas suggest that primer specific to the β-geo insertion in those mice didn't cease or suppress the expression of Grb10 although the total expression level of Grb10 in mutant Grb10KO$^{+/p}$ mice found to be lower than the wild-type mice in both tissue samples.

Although the read depth of the primers specific to Grb10 and β-geo insertion are extremely low in comparison with the total number of reads in each sample suggesting these may be noise.

Detecting of expression from *LacZ* reporter genes shows different expression depending on inserted loci in the gene. Cowley et al. (2014) described two mouse models of Grb10 ablation, which are generated by the integration of a *LacZ* reporter gene-trap cassette in different loci of Grb10KO$^{+/p}$ (Garfield et al., 2011) and Grb10Δ2-4$^{+/p}$ (Charalambous et al., 2003). That shows identical phenotypic consequences, but an expression of the *LacZ* reporter genes was not equivalent as Grb10KO$^{+/p}$ was detectable in the CNS of embryos but Grb10Δ2-4$^{+/p}$ was not. In Adult brain Grb10Δ2-4$^{+/p}$ is detectable but weak relative to Grb10KO$^{+/p}$. Also, mutations in the fruitless gene in Drosophila correlate with the varying behavioural defects as mutations of fru gene leads to either no expressing neuronal cluster or exhibit expression in small numbers of neurons at extremely low levels (Villella et al., 1997; Lee & Hall, 2001).

Galaxy (public web access: http://usegalaxy.org; download: http://getgalaxy.org) is an open, robust web-based documents for data-intensive and transparent computational biomedical research (Blankenberg et al., 2010, Goecks et al., 2010), it manipulates RNA-Seq dataset via the quality filtering steps (Blankenberg et al., 2010) by providing a report for modular set of analyses mainly the basic statistics (Table 3.5 and Table 3.6), per base sequence quality (Figure 3.2 and Figure 3.3) and the adapter content (Figure 3.4 and Figure 3.5). These reports provide a quick overview as an awareness before further analysis. The quality reports showed that all the 60 paired-end reads of the cortex and the subcortical areas samples have a sequence length of 125 bp with zero poor quality sequences, no adapters, and the quality scores in Phred scale for all samples had been above 23 with the majority above 25. The samples can be processed for further computational analysis.

Differentially expressed genes between the wildtype and the mutant Grb10KO$^{+/p}$ samples were identified using the edgeR package supported in R (Robinson et al., 2010). The brain function and structure undergo dramatic changes from embryonic to postnatal development. Detecting highly expressed genes at different stages starting from E18.5 till adult stage may serve as a blueprint for gene expression pattern which gives clues for any behaviour or diseases related to genes during brain development.

The tissue sample at embryonic stage is very small that might result in some contamination in those samples. Therefore, MDS showed some contamination at the stage of E18.5. The visualisation with MDS revealed mutation of Grb10KO$^{+/p}$ cause

no expression differences in the cortex while in the subcortical areas slight difference expression between the WT and mutant Grb10KO$^{+/p}$ at 3M old mouse suggesting dramatic changes at certain tissue during the certain stages. A higher number of differentially expressed genes of both up- and down-regulated showed at age of 3 months old in the subcortical areas of 139 genes and the cortex with 26 genes (Figure 3.7). This result suggests that most transcriptional alterations take place at the stage of 3M old. Heatmap analysis highlights that Grb10 is down-regulated in those stages which confirmed the symptomatic manifestation of social dominance in the mutant Grb10KO$^{+/p}$ mouse model where the behavioural onset appears at adult stages (Garfield et al., 2011). pairwise comparisons between stages revealed that there is no considerable amount of differentially expressed genes shared across the developmental groups in both the subcortical areas (Figure 3.9A) and the cortex (Figure 3.9B and Table 3.8) except 1-3 genes. Grb10 is one of the shared differentially expressed gene at mature stages (3M and 6M) old mice in both the cortex and the subcortical areas. GO enrichment for the overlapped genes in the subcortical areas and the cortex showed no adjusted significant values.

To gain further knowledge into the functional coherence of the differentially expressed genes and recognize their specific targets of biological processes and cellular components, a potential enrichment of specific functional categories among the up- and down-regulated genes was measured using specially written R-based scripts by contrasting the number of annotated genes which obtained from the Ensembl Biomart database version 87 (http://www.ensembl.org/biomart) to a relevant GO with the expected representations of GO terms, with their standard deviations numerically derived from Monte Carlo simulations using 10,000 equally-sized random samples of genes from the list of Ensembl Gene IDs to specify a Z-score with their corresponding p-value, and the p-values, were adjusted for multiple testing using Benjamini-Hochberg correction. Overall GO enrichment clear that the module of down-regulated differentially expressed genes in the subcortical areas targeted the biological process associated with the social dominant behaviour in mutant Grb10KO$^{+/p}$ mice which is occurred at age of 3M old mice where it showed highest number of gene expression (Figure 3.7). These findings and future experiments following may shed light on the neural mechanisms underlying social dominant behaviour in adult mice and more broadly.

Chapter IV. Alternative splicing alterations in the Grb10KO$^{+/p}$

Introduction

Alternative splicing in the brain

The transcriptome is the complete set of transcripts in a cell, and their quantity, for a certain developmental stage or condition. It is essential for interpreting the functional elements of the genome and detecting the molecular constituents of cells and tissues, thus getting knowledge on certain stage development and conditions. Differential splicing or alternative splicing which involves the generation of alternative transcripts exons from different tissues is one of the regular processes by which the exons of primary transcripts (pre-mRNAs) from genes can be spliced in different arrangements during gene expression resulting protein diversity from a single gene coding. It is also taken into consideration to be a key element underlying increased cellular and functional complexity in higher eukaryotes organisms (Matlin et al., 2005; Blencowe, 2006; Ben-Dov et al., 2008; Zhou et al., 2018). Due to its unique part in evolution, as it can increase the rate of evolutionary change in specific exons (Modrek & Christopher, 2003), Nilsson et al. (2010) hypothesised that alternatively spliced transcripts present a new means of regulating gene expression which can probably contribute to protein diversity.

In general, it has been estimated that more than 95% of multiple-exon pre-mRNAs in humans undergo alternative splicing (Pan et al., 2008), also, it has been estimated that 15% of mutations that cause human genetic disease affect splicing (Krawczak et al., 1992). Furthermore, 70-90% of these alternative-splicing events alter the resulting protein products (Kan et al., 2001; Modrek et al., 2001). The comparisons of human and mouse alternative splicing reveal both conserved and species-specific events, with the balance between the two classes depending on the methodology used (Modrek & Lee, 2003; Nurtdinow et al., 2003; Thanaraj et al., 2003).

The brain displays the most complex pattern of alternative splicing compared with other tissues. Several researchers indicated evidence that most frequent alternative exons are neuron specific (Johnson et al., 2003; Modrek et al., 2001; Pan et al., 2008; Watson et al., 2005; Xu et al., 2002; Yeo et al., 2004) as it is most complex in the nervous system throughout vertebrate evolution (Yeo et al., 2004; Chen & Manley, 2009; Barbosa-Morais et al. 2012; Merkin et al., 2012) which might be due to

its contribution to the of brain function, anatomy and development.

Beffert et al. (2005) showed that deletion of a single, activity-dependent alternative exon in the apolipoprotein E receptor 2 (Apoer2) gene, which is important for neuronal cell migration during brain development and for long-term potentiation in adult mice, leads to poor memory and learning tasks. As well, there are many examples of regulated alternative splicing in the brain, which can be concerned in complex processes such as the control of synaptic plasticity associated with cognition and other neural processes (Lipscombe, 2005; Ule and Darnell, 2006).
Recent studies on neocortex suggested as the main site that alternative splicing events that influence cortical development, cell fate, and layering (McKee et al., 2005; Belgard et al., 2011; Zhang et al., 2014; Zhang et al., 2016). Maintaining of the brain-specific alternative splicing program is especially prominent in vertebrate evolution suggesting functionality of spliced products (Barbosa-Morais et al., 2012; Merkin et al., 2012). Dysregulation of this program can lead to neurological disease (Licatalosi & Darnell, 2006). Hence, alternative splicing might be a key mechanism for producing the range of protein activities that support complex brain functions, and targeting certain alternative exons can lead to understanding the regulation of complex biological processes.

Grb10 alternative splicing transcripts

Grb10 protein is expressed as multiple isoforms generated by alternative mRNA splicing (Daly, 1998).

It is transcribed in neurons from a series of downstream alternative promoters, exclusively from the paternal allele (Hikichi et al., 2003; Sanz et al., 2008) which is conserved in the human brain (Arnaud et al., 2003). In mice brain, in the subcortical areas, Grb10 transcribed paternally at age of E14.5 and are maintained in the adult stage while it is excluded from the cortex (Garfield et al., 2011). Although, in situ hybridization for Grb10 in the adult mouse brain has shown that biallelic expression of Grb10 is limited to very few brain regions, suggesting that maternal glial expression in the adult brain is rare (Garfield et al., 2011).

In human, Grb10 gene is located on chromosome 7 and occurs as six isoforms which

are Grb10α (previously referred to as GRB-IR) found mainly in skeletal muscle and pancreas, and in lesser quantities in brain, heart, liver, kidney, lung, kidney and placenta (O'Neill et al., 1996), Grb10β which is most widely and abundantly expressed isoform in most tissues but raised more in skeletal muscle and pancreas (Frantz et al., 1997), Grb10γ which is high in skeletal muscle and in some tumour cells (Dong et al., 1997), Grb10δ, Grb10ε and Grb10ζ (Desbuquois et al., 2013; Morrione, 2003). In contrast, mouse Grb10 gene is located on chromosome 11, it has two main Grb10 splice variants Grb10α (Ooi et al., 1995) and Grb10δ (predominant isoform) (Laviola et al., 1997), in addition, two isoforms are predicted Grb10β1 and Grb10β2 both differ in the alternative splicing of exons 3 and 4 (Philippe et al., 2003). Ensemble transcript identifiers represents 26 transcripts splice variants in human and 9 in mouse (Table 1.1) (Ensembl.org, 2017).

Transcriptome analysis of alternative splicing events

RNA-Seq reads yield insights into the regulation of alternative splicing by way of revealing the use of recognised or unrecognised splice sites in addition to the expression level of exons which are located within highly expressed splicing junctions (Schafer et al., 2015). The technology of Next-generation sequencing leads to detect the mechanisms of genome regulation through sequencing and analysis of the cDNA libraries which are generated from the transcriptome of certain RNA populations in experimental tissues and cells (Kahvejian et al., 2008; Voineagu et al., 2011; Xiao et al., 2012; Xue et al., 2009; Xue et al., 2013). Expressed sequence tags (ESTs) which are derived from fully processed mRNA (after 5′ capping, splicing and polyadenylation) and cDNA sequence databases provide a rich source of information about splicing events occurring in the human and mouse transcriptomes.

The accessibility of large databases of sequenced transcripts and sequenced genomes has provided a good source of information for the analysis and identification of alternative splicing events via using various programs. Thus, Massive databases of alternative splicing events have been established for numerous species such as human, mouse, and rat (Modrek & Lee, 2002; Lee et al., 2003; Thanaraj et al., 2004; Zheng et al., 2005) as well several studies have applied analyses of short cDNA read (mRNA-Seq) data from high-throughput or next generation sequencing technologies to survey alternative splicing in mouse tissues and in human and

mouse cell lines (Bainbridge et al., 2006; Cloonan et al., 2008; Mortazavi et al., 2008, Sultan et al., 2008).

The sequence- and microarray-based analyses have demonstrated that alternative splicing events occur more frequently in transcripts from genes expressed in functionally complex tissues with diverse cell types, for example the brain (Modrek et al., 2001; Xu et al., 2002; Johnson et al., 2003; Yeo et al., 2004; Watson et al., 2005; Pan et al., 2008) and testis (Yeo et al., 2004), or diverse functions as in the immune system (Watson et al., 2005). The ratio between reads including or excluding exons known as percentage spliced in the index (PSI) which is based on two read populations within an RNA-Seq dataset (Wang et al., 2008). It also, suggests how efficiently sequences of interest are spliced into transcripts (Schafer et al., 2015) and the PSI values can be used as fact for comparison of the anticipated occasion quantifications, as it considered to be the ground truth for comparison of the predicted event quantifications (Kahles et al., 2016).

Tools for computational analyses of transcriptional alternative splicing

Understanding the biologic function of alternative splicing has been hindered by the difficulty in systematically identifying and validating transcripts which are coregulated by particular splicing factors (Shin & Manley, 2004). Microarrays using exon-junction probes had been used first for genome-wide investigation in yeast (Clark et al., 2002). In general mRNA-Seq data provide reliable measurements for exon inclusion levels (Pan et al., 2009).

Alternative splicing of transcriptome has been identified, using a variety protocols of analyses, and the availability of large databases of cDNA and expressed sequence tag (EST) sequences enabled large-scale computational studies, which have assessed the scope of alternative splicing in the mammalian transcriptome (Graveley, 2001; Modrek et al., 2001; Clark & Thanaraj, 2002; Kan et al., 2002). Therefore, the availability of extensive transcript sequence data and alternative splicing profiling technologies yields new insights into the nature of the splicing. Researchers follow variable analysis methods related to their research goal and type of selected organism. For instance, the organism such as human or mouse where their genome sequence is available, then identifying transcripts by mapping RNA-Seq reads onto

the sequenced genome should be possible. In contrast, if the sequenced genome of the organism is not available, then quantification would be achieved by using de novo transcriptome assembly method.

To estimate transcript expression and/or gene expression is primarily based on the number of reads that map to each transcript sequence, in spite of the fact that, there are computational tools such as Sailfish (Patro et al., 2014) that rely on k-mer counting in reads without the need for mapping. However, the simplest way for quantification is to sum raw counts of mapped reads using programs such as featureCounts (Liao et al., 2014) or HTSeq-count (Anders et al., 2015). In order to handle the problem of any related transcripts of reads, many sophisticated algorithms were developed. Example of Algorithms that quantify expression from transcriptome mappings includes eXpress (Roberts & Pachter, 2013), RSEM (Li & Dewey, 2011), Sailfish (Anders et al., 2015) and Kallisto (Bray et al., 2016) among others.

Several other bioinformatics analyses methods of RNA-Seq data can be used for Identification of novel transcripts such as Cufflinks (Roberts et al., 2011), iReckon (Mezlini et al., 2013), StringTie (Pertea et al., 2015), SLIDE (Li et al., 2011), Montebello (Hiller & Wong, 2013). Some methods have specific algorithms for alternative splicing analysis which mainly based on two categories. The first category integrates isoform expression estimation with the detection of differential expression to reveal changes in the proportion of each isoform within the total gene expression such as CuffDiff2 (Trapnell et al., 2013). The flow difference metric (FDM) (Singh et al., 2011). rSeqDiff (Shi & Jiang, 2013).

The second category showed greater accuracy as it skips the estimation of isoform expression and detects signals of alternative splicing by comparing the distributions of reads on exons and junctions of the genes between the compared samples. The examples for this category are DEXseq (Anders et al., 2012) and DSGSeq (Wang et al., 2012). rMATS (Shen et al., 2014). rDiff (Drewe et al., 2013). DiffSplice (Hu et al., 2013) and SplAdder (Kahles et al., 2016).

Aim

Our objective is to determine the common alternative splicing transcripts in the neocortex and the subcortical areas at different development stages related to the Grb10KO$^{+/p}$ compared to WT.

Methods

Analyses of alternative splicing transcripts

RNA-Seq is a recently developed approach to transcriptome profiling that uses sequencing technologies. It provides a far more precise measurement of levels of transcripts and their isoforms. To check the alternative splicing events across mouse brain in the subcortical areas and the cortex at different developmental stages, two different softwares were employed. The first software employed was SplAdder software (Kahles et al., 2016). The reads from RNA-Seq experiments were first mapped to the Mus musculus reference genome (release 87) using the HISAT2 (version 2.0.3) software (http://ccb.jhu.edu/software/hisat2). Read alignments were then processed with the SAMtools software package (Li et al., 2009) to generate position sorted bam and bai file for each sample. For alternative splicing analysis, I used SplAdder software (Kahles et al., 2016) and then for each sample the percent spliced index (PSI) values from annotated exon_skip_merge_graphs_C3. confirmed file was selected, which indicates the efficiency of splicing a specific exon into the transcript population of a gene. With PSI values of each sample, the differential expression analysis between WT and mutant Grb10KO$^{+/p}$ mice for each developmental stage in the subcortical areas and then the cortex was conducted using edgeR tool which supported in R (Robinson et al., 2010).

The second software employed was Kallisto (Bray et al., 2016). Analysis using Kallisto was obtained as mentioned in chapter III. The generated quantification abundances of all annotated transcripts of the RNA-Seq for each sample was then used. The edgeR tool supported in R (Robinson et al., 2010) was used for differential expression analysis between WT and mutant Grb10KO$^{+/p}$ mice at the

different developmental stage, in the subcortical areas and then the cortex.

Results

In order to identify alternative splicing events in the wild-type and KO mice, I used two alternative pipelines. The SplAdder software generates several different output files, mainly 5 events that have been detected by SplAdder namely: merge_graphs_alt_3prime_C3. confirmed, merge_graphs_alt_5prime_C3. confirmed, merge_graphs_exon_skip_C3. confirmed, merge_graphs_intron_retention_C3. confirmed, and merge_graphs_mult_exon_skip_C3. confirmed.

For each event type, 4 separate set of output files is also generated for an efficient query or/and further analysis. For our purpose, I have chosen for each sample their PSI values which considered ground truth for comparison of the predicted event quantifications (Kahles et al., 2016) in merge_graphs_exon_skip_C3. confirmed file. I found that in subcortical areas there are 5132 alternative splice events, with a total of 1548 genes within it 4983 alternative splice events and 209 genes differ from the cortex while 149 alternative splice events and 1339 genes are shared (Table 4.1 and Figure 4.1). The alternative splice within a gene ranged between 1-33 events (Table 4.1). In contrast, the cortex showed 5229 alternative splice events, with a total of 1540 genes in which within it 5080 alternative splice events and 201 genes differ from the subcortical areas (Table 4.1 and Figure 4.1), while the alternative splicing events within a gene ranged between 1-37 (Table 4.1).
Also checking the total number of alternative splicing events, number of genes and quantity of alternative splicing within a gene in the subcortical areas and the cortex in each of the five developmental stages revealed very slight difference in each of those stages in both brain tissue (Table 4.2) The highest alternative splicing events occurred in the cortex at 6M (=6024), while the lowest occurred in the subcortical areas at E18.5 (=5633). Both the highest (=1717) and the lowest (=1667) number of genes appeared in the subcortical areas at 1W and E18.5, respectively. Lastly, the highest number of alternative splicing within a gene appeared in the cortex at 6M old mice (=43) (Table 4.2).

Differential expression analyses based on transcripts, using edgeR software showed

no adjusted significant level of differential expression i.e. FDR-Adjusted P-value < 0.05 between the WT and mutant Grb10KO$^{+/p}$ mice at all the developmental stages. Surprisingly Grb10 isoforms were not detected by any of the SplAdder generated output files although several have been reported previously in the literature.

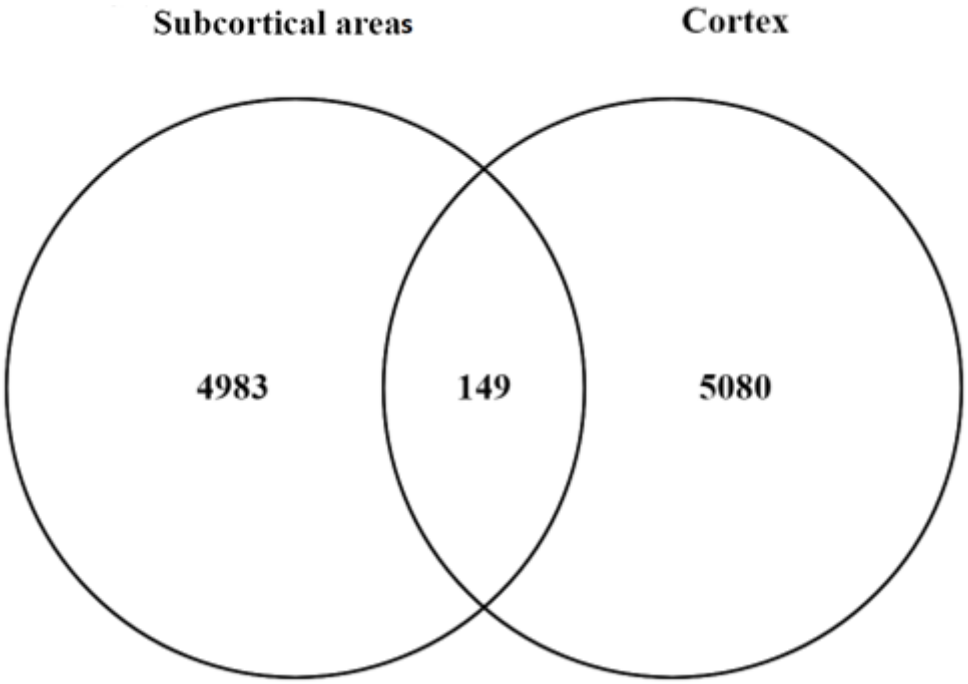
Table 4.1. Alternative Splicing using SplAdder. Total number of alternative splicing events, number of genes and quantity of alternative splicing within a gene in the subcortical areas and the cortex generated by SplAdder.

Differences	Subcortical areas	Cortex
Total number of alternative splicing events.	5132	5229
Total number of genes.	1548	1540
Number of alternative splicing within a gene.	1-33	1-37

Table 4.2. Alternative Splicing using SplAdder. Total number of alternative splicing events, number of genes and quantity of alternative splicing within a gene in the subcortical areas and the cortex in each of the five developmental stages generated by SplAdder.

Differences	Subcortical areas					Cortex				
	E18.5	**1W**	**1M**	**3M**	**6M**	**E18.5**	**1W**	**1M**	**3M**	**6M**
Total number of alternative splicing events.	5633	5872	5870	5870	5875	5867	5762	5980	5978	6024
Total number of genes.	1667	1717	1701	1704	1712	1702	1677	1705	1706	1709
Number of alternative splicing within a gene.	1-33	1-34	1-33	1-33	1-33	1-41	1-38	1-42	1-42	1-43

A. Shared alternative splicing events



B. Shared genes

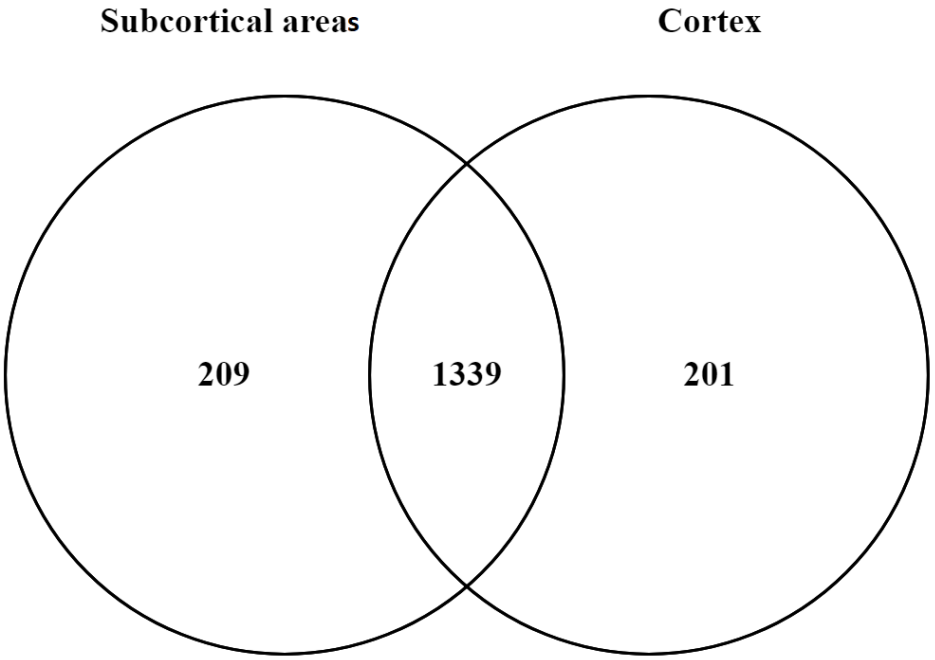


Figure 4.1. Overlapped alternative splice variants/genes between the subcortical areas and the cortex. Number of shared alternative splice (A) and shared genes (B) between the subcortical areas and the cortex.

The second software used to identify alternative transcripts was Kallisto (which was also used to carry out transcriptome annotation in the previous chapter) Output file of abundances.tsv which is generated by Kallisto were analysed. This file generates all transcripts from the RNA-Seq reads. Differential expression analysis comparing the wild-type and the mutant Grb10KO$^{+/p}$ transcripts samples using the edgeR package supported in R (Robinson et al., 2010) identified a total of 103215 transcripts (Table 3.7). Splice variants of Grb10 transcript appear at different developmental stages in both brain tissue and only 3 Grb10 transcript splice variants have been detected out of 9 common splice variants (Table 4.3). These are ENSMUST00000148254 (mGrb10-208) which consist of 351 bp, it has four exons and it is a non-coding transcript, ENSMUST00000093321 (mGrb10-201) which consist of 5061 bp, it has 17 exons and its protein contains 596 amino acids, and ENSMUST00000109654 (mGrb10-203) which consist of 4755 bp, it has 16 exons and its protein contains 541 amino acids (Figure 4.2) (Ensembl.org, 2017).

In the subcortical areas the adjusted significant differential expression (FDR < 0.05) result showed the highest number of gene transcripts at 3 months old mice (=256 transcripts), while the lowest number at 1week old mice (=115 transcripts) (Table 4.3). The most common splice variant of Grb10 transcript with FDR < 0.05 is mGrb10-208, it is detected at all the 5 developmental stages. However, at 3 months old mice 2 other splice variants have been detected which are mGrb10-201, mGrb10-203, while at 6 months old mice besides the mGrb10-208 splice variant, mGrb10-201 has been also detected (Table 4.3).

In the cortex, the adjusted significant differential expression (FDR) result showed the highest number of gene transcripts at 1month old mice (=223 transcripts), while the lowest number at 6 months old mice (=139 transcripts) (Table 4.3). Only 3 months and 6 months old mice show splice variant of Grb10 transcript which is also mGrb10-208, in addition, 6 months old mice also show mGrb10-201 Grb10 transcript splice variant.

Table 4.3. The edgeR result of transcripts using Kallisto's output. A total number of differentially expressed transcripts (FDR < 0.05), and detected Grb10 splice variants in the subcortical areas and the cortex at different developmental stages of the mice obtained by Kallisto, ENSMUST00000148254 (mGrb10-208), ENSMUST00000093321 (mGrb10-201), ENSMUST00000109654 (mGrb10-203).

Tissue.	Subcortical areas					Cortex				
Developmental stages.	E18.5	1W	1M	3M	6M	E18.5	1W	1M	3M	6M
Adjusted significance level of a total number of Transcripts.	214	115	133	256	139	146	183	223	141	139
Grb10 splice variants.	mGrb10-208	mGrb10-208	mGrb10-208	mGrb10-208, mGrb10-201& mGrb10-203	mGrb10-208& mGrb10-201				mGrb10-208	mGrb10-208& mGrb10-201

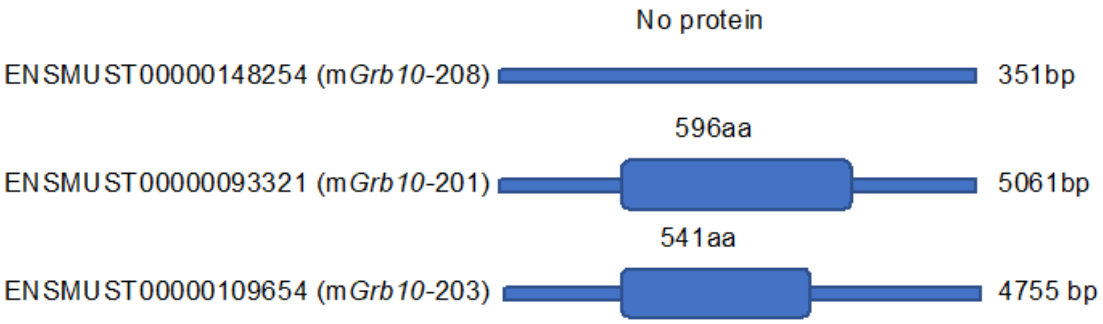


Figure 4.2. Detected Grb10 transcript splice. A schematic of the three major transcript splice variants that have been detected in the brain tissue sample.

Discussion

Alternative splicing of the transcriptome is widespread in mammalian gene expression, and variant splice patterns are often specific to different developmental stages, specific tissues or a disease or a behaviour condition. The potential to identify and characterize the regulated alternative splicing events and alternative variant

transcripts will provide an entirely new view of how different processes and regulation of the cells. In humans, more than half of the genes show alternative splicing (Lander et al., 2001; Modrek et al., 2001; Johnson et al., 2003), and many mRNA isoforms can be linked to specific diseases (Cooper & Mattox, 1997; Beck et al., 1999; Garcia-Blanco et al., 2004). Therefore, understanding the regulation mechanism of alternative splicing and mRNA isoforms is a major goal of the researcher.

My main aim in this chapter was to obtain an inventory of all genes and their alternative splicing events along with their estimated expression levels in the cortex and the subcortical areas at different developmental stages. The SplAdder software has the ability to predict splicing changes and the quantification events, which has been used to determine whether difference occurs in the proportions of these alternative splicing events across mice brain tissues at different developmental stages. I found that SplAdder software shows no big differences at a total number of alternative splicing events, the total number of genes and the number of alternative splicing events within a gene between the subcortical areas and the cortex (Table 4.1). Although the highest alternative splicing events found in the cortex at age of 6 months (=6024) (Table 4.2) this is because in developing cortex, inclusion of the exon increases gradually (Weyn-Vanhentenryck et al., 2018). Also, there has been no adjusted significance level of differential expression at all the developmental stages. Surprisingly, the SplAdder output didn't show alternative splicing events of Grb10. Precise detection of the exons location in each gene is an important measure of the performance of gene identification algorithms (Reese et al., 2000).

Difficulty in distinguishing Grb10 alternative splicing events from RNA-Seq data may be due to a different algorithm that SplAdder tools used as different methods perform best in different data sets or sources of error (Rapaport et al., 2013; Soneson & Delorenzi, 2013). Failure to detect alternatively spliced forms of Grb10 may be due to that some alternatively spliced mRNA forms are miscategorised as genomic DNA, causing to be excluded by the procedure for analysing its splicing. The edgeR result with Kallisto output showed that the highest number of significantly differentially expressed (FDR < 0.05) splice variant transcripts appeared in the subcortical areas at 3 months old mice with a total number of 256 (Table 4.3). This is may be due to the highest number of gene expression appeared at 3 months old mice in the subcortical areas (Figure 3.7), supporting the idea that tissue-regulated splice variant transcripts plays important roles in the differentiation of the tissue at certain developmental

stages leading to vast regulations of mRNA at that stage which may relate to the unusual patterns of splicing-factor expression observed in the mature brain of the mouse, suggesting aspects of developmental regulation of splice variant at the tissue level.

Transcriptome level appears to have different regional distributions in the brain tissue and at different developmental stages (Table 4.3), these differences reflect developmental and functional differences between brain regions. This result also can be supported by the transcriptomic study of brain microvessels in neonatal and adult mice, showing the difference in expression level where adult brain exhibited the highest mRNA expression levels for most genes (Porte et al., 2017). Another good example is given by Han et al. (2009) who found substantial transcriptome splicing variation between 2 developmental stages (embryonic day 18 and postnatal day 7) of the same tissue of the brain (cortex) in the mouse. Furthermore, in the rat, whole transcriptome sequencing of the brain reveals changes in transcript expression, isoform usage, and non-coding RNAs with age (Wood et al., 2013). In addition to that, Mouse CNS transcriptomes responded to age, energy intake and gender in a regionally distinctive manner (Xu et al., 2007). This expression information suggested critical functions for some of the regulatory genes and provided solid confirmation for a highly dynamic transcriptome during mouse brain development.

Grb10 interacts differentially with different receptor (He et al., 1998). Absence of certain domain of Grb10 can impact on its interaction as a study done by He et al. (1998) showed that isolated BPS domain interacted less well (~50%) with the IR cytoplasmic domain than did BPS and SH2. This could be due to difficulty in folding of the BPS to interact with the receptor in the absence of the SH2 domain. Therefore, it is important to know the tissue-specific expression of the transcript splice variants for understanding the social dominant roles of Grb10 in the central nervous system. I found that in the subcortical areas where mutation of Grb10 leads to social dominant behaviour, the most predominant differentially expressed splice variant of Grb10 transcript is mGrb10-208 which doesn't translate into protein due to the lack of an open reading frame (ORF). It is transcribed at all the 5 developmental stages, while at 3 months old in addition of this splice variant two more are represented which are mGrb10-201 and mGrb10-203. The last transcript splice variant is distinguished specifically at this stage. This might suggest that specific splice variants are characteristic of specific tissue/cell at a specific stage for a specific condition.

In contrast, the presence of mGrb10-208 at only two developmental stages (3M & 6M old mice) in the cortex confirmed that expression of the paternally inherited allele of Grb10 is restricted to most subcortical areas which continue into adulthood (Garfield et al., 2011). Presence of mGrb10-201 at 6 months old mice in both the subcortical areas and the cortex suggest unusual patterns of splice variant expression in the mature brain of the mouse.

Defining the set of alternative splicing events conserved in mouse is primary interest in efforts to understand the biological importance of the social behaviour of splicing regulation but the major goal will be to integrate this result into knowledge of coordinated gene expression programs and networks that may be controlled at other levels in addition to transcription which may require further study.

Chapter V. Changes in the regulatory architecture of Grb10KO$^{+/p}$ mice

Introduction

Gene Co-expression network analysis

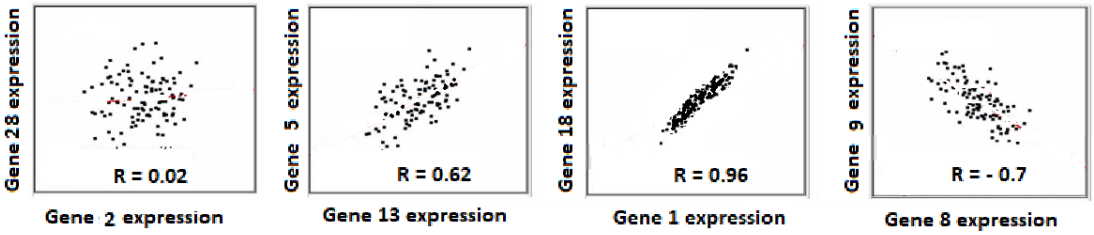
Gene expression levels can be used to assess the relevance of certain groups of genes for a cell, group of cells or tissue at a specific time point. Differential gene expression analyses comparing two conditions or genetic backgrounds can help us to better understand the molecular functions of specific genes. The development of high-throughput technologies such as microarrays and RNA sequencing along with the data analysis methods, makes it easier to identify the gene function from a genome-wide perspective (Carpenter & Sabatini, 2004; van Dam et al., 2015).

Proteins act in concert to carry out cellular functions, communicate with other cells and react to the changing extracellular environment. In recent years, the interaction and between molecular pathways and the plasticity of molecular cascades have been better recognised (Monzon-Sandoval et al., 2016). Genes are increasingly understood as part of regulatory networks where groups of genes contributing to specific molecular functions are up- or down-regulated or even reorganised depending on cell's needs. As transcriptome profiling costs have decreased allowing an increase in the number of biological replicas assessed, gene network analyses based on assessment of co-expression have become possible (Oldham et al., 2006). Co-expression gene networks are constructed based on similarity in gene expression profiles (co-expression) in all possible gene pairs. This allows identification of which genes tend to demonstrate a coordinated expression pattern over a group of samples. In the resulting gene networks, every node symbolizes a gene and each edge represents the presence and the strength of the co-expression relationship.

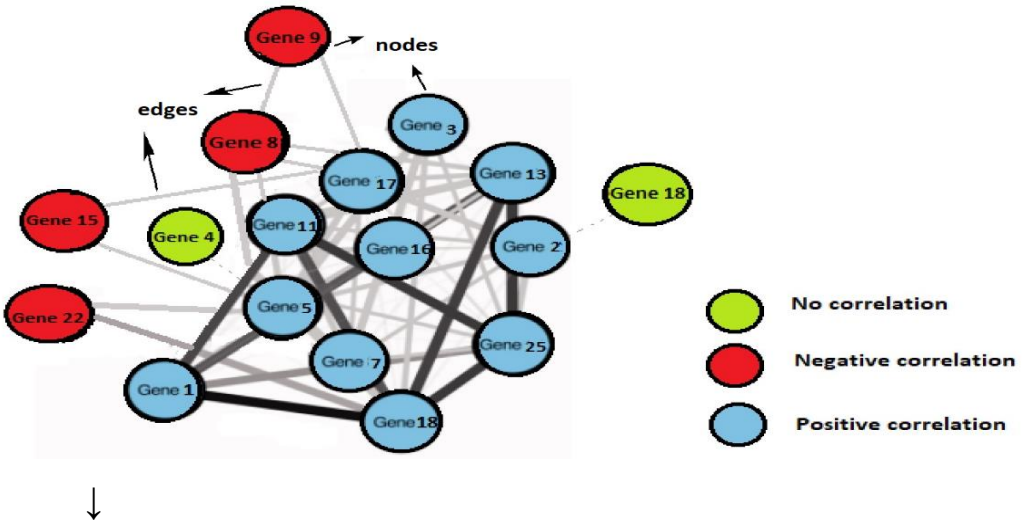
Weighted gene co-expression network analysis (WGCNA) is one of the methods used to characterise gene function as well as gene-disease predictions from wide gene expression of a genome. Highly interconnected genes within the network have been found to share biological modules or pathways demonstrating an effectiveness at assigning putative functions to genes based on the functional annotation of their co-expressed partners leading of improving the understanding of regulatory networks (Stuart et al., 2003; Han et al., 2004). Thus, by reconstructing co-expression networks we can associate genes of unknown function with biological processes (van Dam et al., 2017). WGCNA can also provide information on the functional interactions among

groups of genes and fundamental regulatory architecture associated to global expression profile as co-expressed genes are possibly under the concerted control of a common complement of transcriptional regulators (Yu et al., 2003; Marco et al., 2009; Gaiteri et al., 2014). In general, WGCNA involves using expression data profiles and then calculating pairwise correlation for each possible gene pair of genes (Figure 5.1). These pairwise correlations then represented as a hierarchal network where each gene is represented as a node and correlations above a threshold are denoted as lines. Clustering analysis is used to defined modules within the network. Modules can then be characterised by establishing enrichment in cellular function annotation.

A.



B.



C.

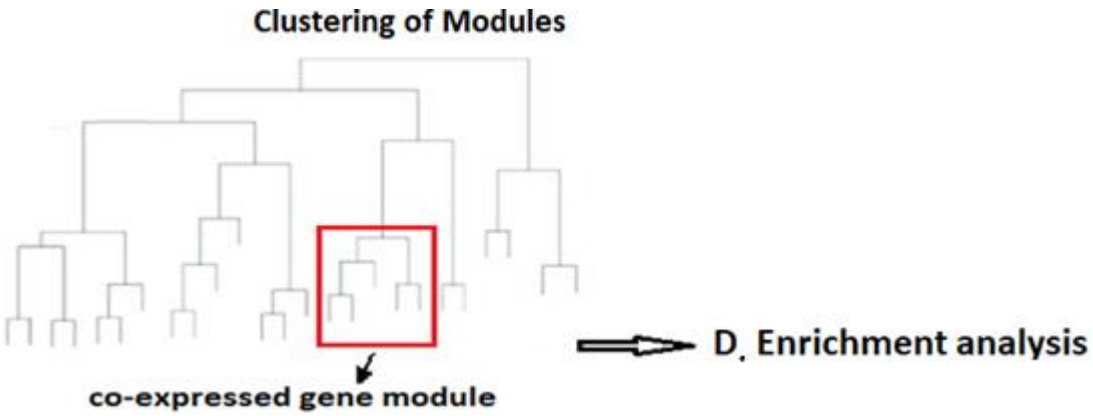


Figure 5.1. Steps of co-expression network analysis. A) each possible pairwise correlation of gene expression is determined, B) These pairwise correlations can then be represented as a network, C) clustering analysis used to define modules within the network, D) functional enrichment identified from those modules.

Differential Co-expression network analysis

Differential co-expression analysis refers to the comparison of two or more gene co-expression networks reconstructed from samples in different conditions. Such analyses can help to identify potential changes in the gene network regulatory architecture (de la Fuente, 2010; Gaiteri et al., 2014; Monzon-Sandoval et al., 2016). It is thought that genes that are differentially co-expressed between groups of sample are more likely to be regulators and potentially more relevant to understanding phenotypic variation (Kostka & Spang, 2004; Hu et al., 2009; Hudson et al., 2009; Amar et al., 2013). Past studies have used differential co-expression network analyses to identify genes underlying differences between different species (Gao et al., 2012; Monaco et al., 2015), cellular changes (Choi et al., 2005), cell types (Riquleme & Lubovac-Pilav, 2016; Zeisel et al., 2015), tissues (Pierson, 2015; Reznik & Sander, 2015), healthy and disease samples (Amar et al., 2013; Hu et al., 2009; Hudson et al., 2009; Kostka & Spang, 2004) or developmental states (Monzon-Sandoval et al., 2016).

Several programs have been developed to assess differential clustering analysis, and the common ones are Weighted Gene Co-expression Network Analysis (WGCNA), DiffCoEx and Differential Correlation in Expression for meta-module Recovery (DICER), all of which first identify modules co-expressed across the full set of study

samples (van Dam et al., 2017). GCNA is widely used, it performs well under different conditions and purposes (Langfelder & Horvath, 2008) it determines the activity and importance of each module in each subpopulation of samples (van Dam et al., 2017), DiffCoEx (Tesson et al., 2010) focuses on modules that are differentially co-expressed with the same sets of genes as it groups genes based on their differential co-expression behaviour (van Dam et al., 2017), DICER (Amar et al., 2013) may be particularly useful for time series experiments in which co-expression changes are gradual (van Dam et al., 2017).

Co-expression in the brain

Changes in gene expression profiles throughout various developmental stages are controlled by a fundamental network of regulatory interactions between individual genes. In human, the development of the central nervous system has an intricate pattern regulated process that appears over a prolonged period of time in response to transmuting needs for a variety of cellular functions relying on a strict temporal and regional coordination of complex patterns of gene expression (Monzón-Sandoval et al., 2016). High variation in gene expression is especially pronounced at a specific stage of development. Stead et al. (2006) stated that in the rat brain the spectacular changes in the level of gene expression for most genes occur at the early stage of postpartum life at about 1-2 weeks and plateau thereafter.

Throughout the development, changes in gene expression profiles are underlined by a fundamental network of precise regulatory interactions between individual genes (Neph et al., 2012). The assembly complexity of genetic components and molecular result majority of the cellular processes (Hartwell et al., 1999) depend on stable regulatory interactions between groups of genes throughout development. Furthermore, many of genes have the potential to participate in many separate and sometimes unrelated biological functions (Harris et al., 2004) intimating the presence of causal events of regulatory reassembly that leads to new functional associations.

Genes which are linked by regulatory interactions due to their functional association tend to exhibit same expression patterns (Eisen et al., 1998; Homousz & Kudlicki 2013). This coordinated expression can be determined by observing the correlations in expression levels between groups of genes of suitable samples. Analysis of

clustering based on co-expression patterns used to describe modules or groups of correlated genes which may from pathways, molecular complexes, common signalling and regulatory circuits (Oldham et al., 2006; Oldham et al., 2008; Saris et al., 2009; Usadel et al., 2009; Torkamani et al., 2010; Obayashi & Kinoshita, 2011; Zhang et al., 2012A).

Co-expression of Grb10

Grb10 is an imprinted gene expressed from the paternal allele in the brain from foetal life into adulthood and ablation of this expression causes increased social dominance (Garfield et al., 2011). Several studies showed biological regulation of Grb10 co-expression, in the regulation of apoptosis Nantel et al. (1998) reported that expression mutants of Grb10 induced apoptosis of the cell while the co-expression of wild-type Grb10 increases cell survival.

In yeast two-hybrid system co-expression of LexA DNA-binding domain fused to the intracellular portion of the IGF-I receptor (LexA-IGFIR beta) with an activated domain of Grb10 (Ad-Grb10) resulted in strong activation of two genes namely: *LacZ* and LEU2 reporter genes (Dey et al., 1996). Nantel et al. (1998) stated that Raf1 and MEK1 signalling pathway show co-expression with the SH2 domain of the Grb10 adapter protein, while Murdaca et al. (2004) showed that Grb10 acts as a positive regulator in VEGF-R2 (Vascular endothelial growth factor receptor-2) signalling and protects it from degradation by co-expression with Nedd4 (neural precursor cell expressed developmentally down-regulated protein). Also, Vélez et al. (2016) established that Nedd4-2 expression increased when Grb10 adaptor was co-expressed with the channel in HEK293 cells. Jahn et al. (2002) proved that co-expression of Grb10 and c-kit activate Akt in a synergistic way. Colley et al. (2009) stated that during neurotrophic signalling pathway both nShc (neuronal Src homology and collagen) and Grb10 adaptor protein expression is reduced when co-expressed with Kv1.3 (Shaker voltage-gated potassium channel). Cook and Fadool (2002) also reported that in HEK 293 cells (Human Embryonic Kidney 293 cells) the co-expression of Kv1.3 with Grb10 causes a decrease in Kv1.3 tyrosine phosphorylation. Grb10 participate in the regulation of cardiac potassium (K+) channels by co-expression with KCNQ1-KCNE1, hERG and other Kv (voltage-gated potassium) channel leading to downregulation of these channels (Ureche et al., 2009).

Co-expression gene network based analysis has yet to be carried out to uncover potential molecular interactors of Grb10 in any tissue or developmental stage. Carrying out a co-expression gene network could shed light on the function and molecular interactors of Grb10 in the brain.

Aim

To plan to carry out differentially co-expressed genes network analysis to detect any relation between sets of genes and gene regulatory networks in the subcortical areas and the cortex triggered by a mutation of Grb10KO$^{+/p}$ which might provide valuable insight into the genetic basis for understanding the social dominance behaviour.

Methods

Statistical analysis of co-expression

All the measurable examinations were completed in R. The Differential co-articulation examination content was acquired from Tesson et al. (2010) which is already carried out in R. The result of all gene expression counts for all the 60 samples of the WT and mutant Grb10KO$^{+/p}$ mice at the 5 developmental stages (E18.5, 1W, 1M, 3M and 6M old) in both the subcortical areas (30 samples that includes 15 WT and 15 mutant Grb10KO$^{+/p}$ mice) and the cortex (30 samples that includes 15 WT and 15 mutant Grb10KO$^{+/p}$ mice) obtained from Kallisto (referred in chapter III) have been used. All genes displaying zero variance across samples were removed from the analysis. Protein-coding genes were selected only and downloaded from Ensembl version 80 annotations (http://www.ensembl.org/index.html).

Co-expression structure clustering analysis

Similarity of expression profiles across all the samples between the WT and mutant Grb10KO$^{+/p}$ mice was assessed in subcortical and neocortex areas. Pearson

correlation coefficient was obtained between the normalized average expression values for each gene of all possible pairs of expression profiles. Co-expression matrices (defined as the Pearson correlation matrix between all possible pairs of genes) were obtained for both brain regions pooling all samples across developmental stages in WT and mutant Grb10KO$^{+/p}$.

Differential Co-expression network analysis

To quantify changes in the global pattern of co-expression between the conditions (WT and mutant Grb10KO$^{+/p}$) at different developmental stages, differential co-expression analysis was performed as described by Tesson et al. (2010) which builds on the commonly used WGCNA framework for co-expression analysis by calculating correlation coefficients for all possible gene pairs separately for the WT and mutant Grb10KO$^{+/p}$, resulting a correlation matrix for each stage. Then the adjacency difference matrix between the conditions using the soft threshold parameter beta =6 was computed to achieve a scale-free degree distribution with fitting index R^2 >=0.80. Hierarchical clustering was performed using the Topological Overlap of the adjacency difference as input distance matrix. The dynamic tree cut function (implemented in R) was used to identify gene modules with a minimum cluster size of 100 genes. Each module is assigned a colour and the modules were merged when the module's eigengenes correlation was higher than r =0.8.

In Addition to that, the co-expression between Grb10 and all the remaining genes was calculated in the subcortical areas corresponding to WT mice. The top 30 co-expressed genes (including Grb10) with a correlation value of 0.8 or higher then selected. The co-expression among these 30 genes was calculated both for WT and mutant mice samples and a co-expression network for the subcortical areas was generated with the R package 'igraph' taking into account both conditions WT and mutant. The same set of genes were used to calculate and generate the co-expression values in the cortex and its network.

Bar plots and Heatmap

Bar graph was used in the *R* software to represent the individual size of the cluster

(module) represented in the cortex and the subcortical areas. Heatmap package endured in the *R* software proposed by Ploner A. (2015) was performed to portray the graphical portrayal of the modules in the cortex and the subcortical regions. Significance permutation analysis and Adjusted p-value (Bonferroni)

This function computes the dispersion value that quantifies the change in correlation between two conditions to determine the statistical significance of the observed correlation changes in each module, I performed a permutation analysis as described and implemented by Tesson et al. (2010) where 1000 permutations are executed on the expression values per module and the proportion of changes in correlation higher than the one observed is determined (Tesson et al., 2010). Finally, the p-value calculated from the data is then compared to this distribution to determine an empirical adjusted p-value where the adjustment method of the Bonferroni correction ("Bonferroni") was used in R to run the adjusted p-value in each of those cluster modules in the subcortical areas and the cortex.

Gene Ontology enrichment analysis

Gene ontology biological process (GO) annotation was downloaded from Ensembl Biomart version 87 (http://www.ensembl.org/index.html) and chose just those GO terms containing 10 categories with at least 2 genes for which expression data was available. Enrichment analysis for each of the modules identified through differential co-expression analysis was carried out and the threshold of 10 genes related to the GO term was made as described earlier (referred in chapter III) concisely; statistically significant overrepresentation of GO terms was assessed based on a Z-score test. Mean and standard deviation for the expected number of genes annotated to each GO term per module was estimated based on 10,000 equally-sized random samples drawn from the background gene population. P-values were adjusted for multiple testing using Benjamini-Hochberg correction and GO enrichments were regarded significant when FDR < 0.05.

Furthermore, GO enrichment analysis was carried out for the 30 co-expressed genes (including Grb10) with correlation value of 0.8 or higher and which were used to calculate and generate the co-expression network in the WT and its corresponding in the mutant in the subcortical areas and the cortex.

Results

Significant changes in regulatory architecture associated with the Grb10KO$^{+/p}$

Examining of all gene expression counts for all the 60 samples of the WT and mutant Grb10KO$^{+/p}$ mice at the 5 developmental stages (E18.5, 1W, 1M, 3M and 6M old) in both the subcortical areas (30 samples that includes 15 WT and 15 mutant Grb10KO$^{+/p}$ mice) and the cortex (30 samples that include 15 WT and 15 mutant Grb10KO$^{+/p}$ mice) obtained from Kallisto resulting in a total of 22114 genes after removing from the analysis all genes displaying zero variance across the samples in order to maximize the number of genes included.

Differential co-expression network dendrogram and gene modules in the subcortical areas.

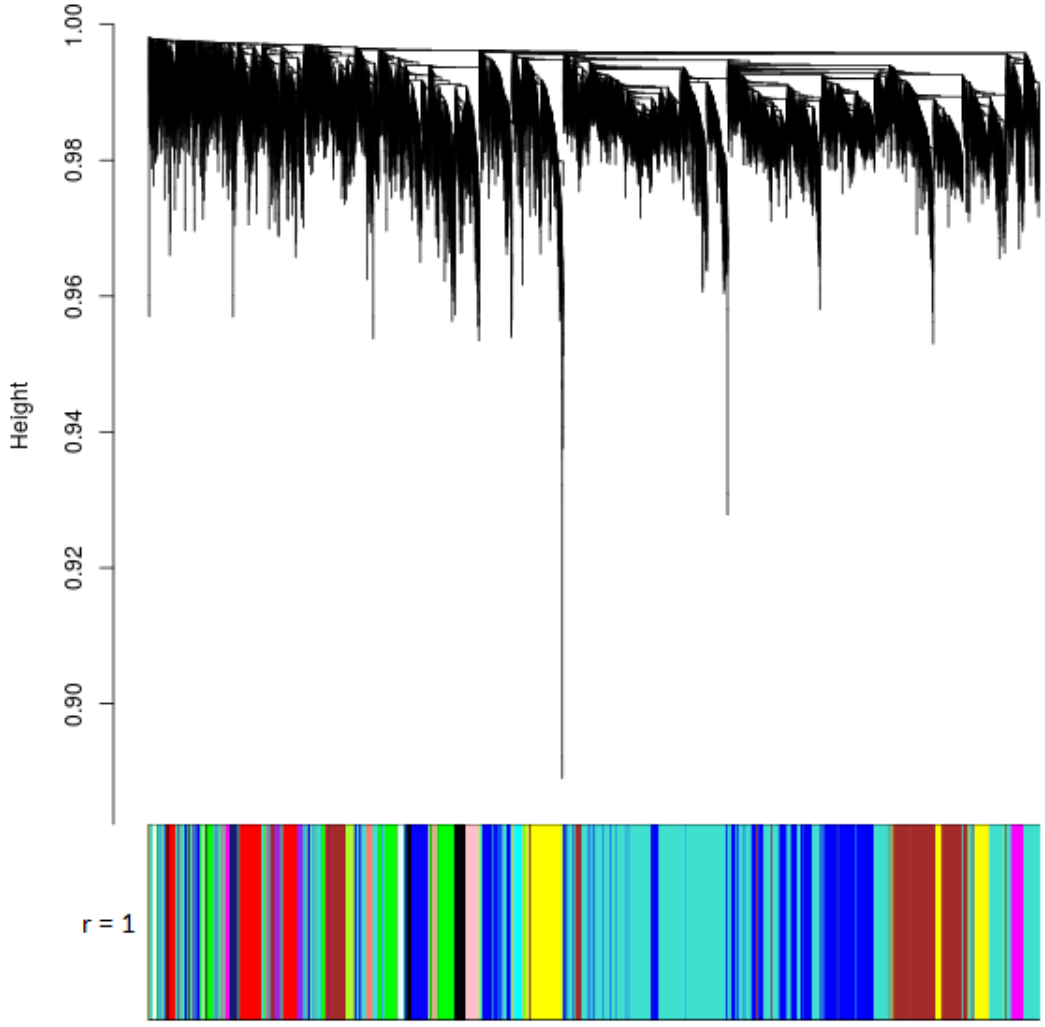


Figure 5.2. Differential co-expression network dendrogram and gene modules. Dendrogram showing hierarchical clustering based on the topological overlap of the adjacency matrix based on co-expression differences (see methods) in the subcortical areas and the corresponding gene modules indicated in different colours underneath.

To characterize the pattern of regulatory changes occurring during the developmental stages, differential co-expression analysis was conducted as described by Tesson et al. (2010) in the subcortical areas and the cortex samples. This method groups genes together when their correlations with the same sets of genes change between different conditions. Hierarchical clustering used to describe the strength of the relationship between all genes in the data, where it identifies modules of differentially co-expressed genes in the subcortical areas (Figure 5.2) and the cortex (Figure 5.3).

Differential co-expression network dendrogram and gene modules in the cortex.

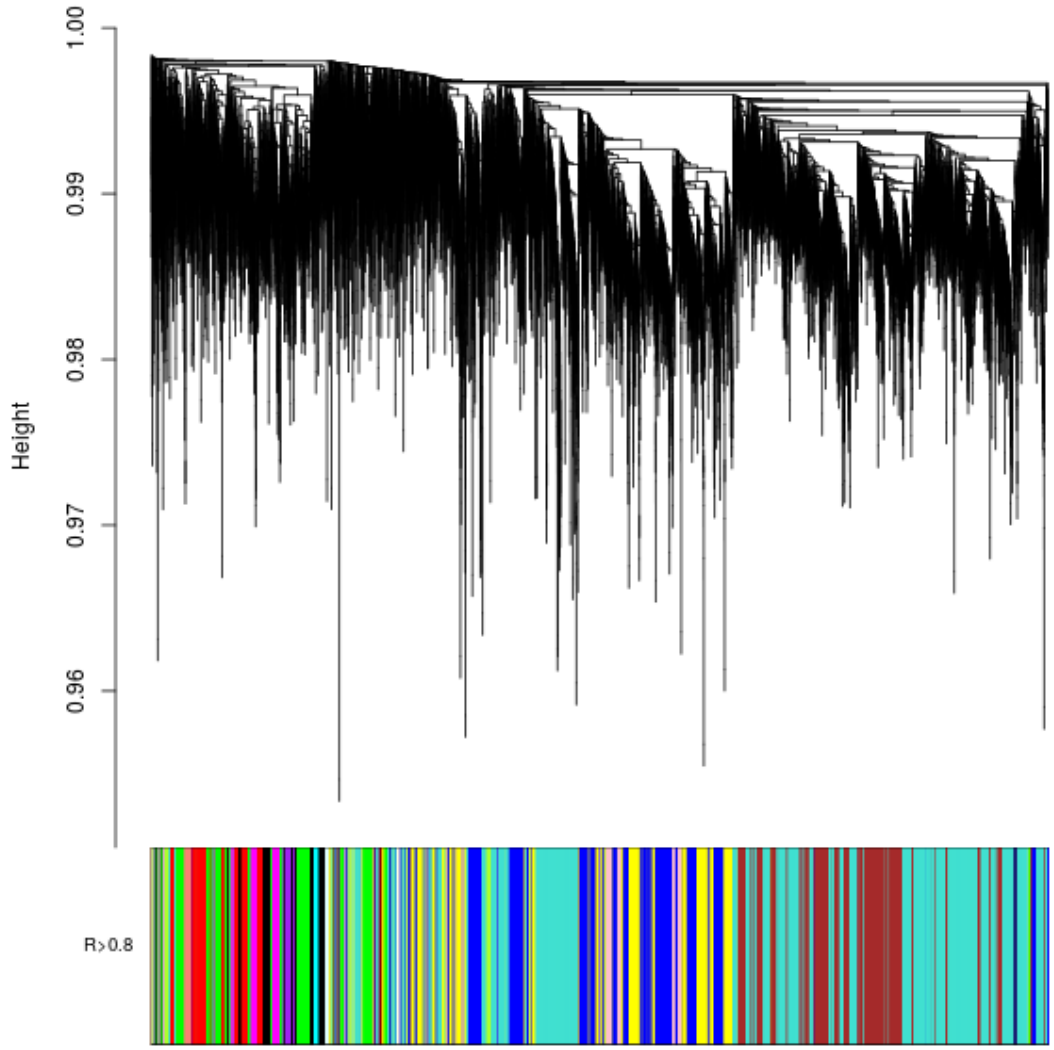


Figure 5.3. Differential co-expression network dendrogram and gene modules. Dendrogram showing hierarchical clustering based on the topological overlap of the adjacency matrix based on co-expression differences (see methods) in the cortex and the corresponding gene modules indicated in different colours underneath.

The clustering of the subcortical areas was merged when their eigengene correlations were high (R > 0.8). The same procedure repeated for cortex resulting in a total of 22 differential co-expression modules ranging in size from 50 to 6353 genes (Figure 5.4). Subcortical areas showed a similar number of differential co-expression modules (22) ranging in size from 53 to 5636 genes (Figure 5.5). However, the module where the Grb10 is presented in the subcortical areas (dark red module) showed less number of genes (52) compared with the module (yellow module) in the cortex which showed 2289 genes.

Differential co-expression gene modules in the subcortical areas.

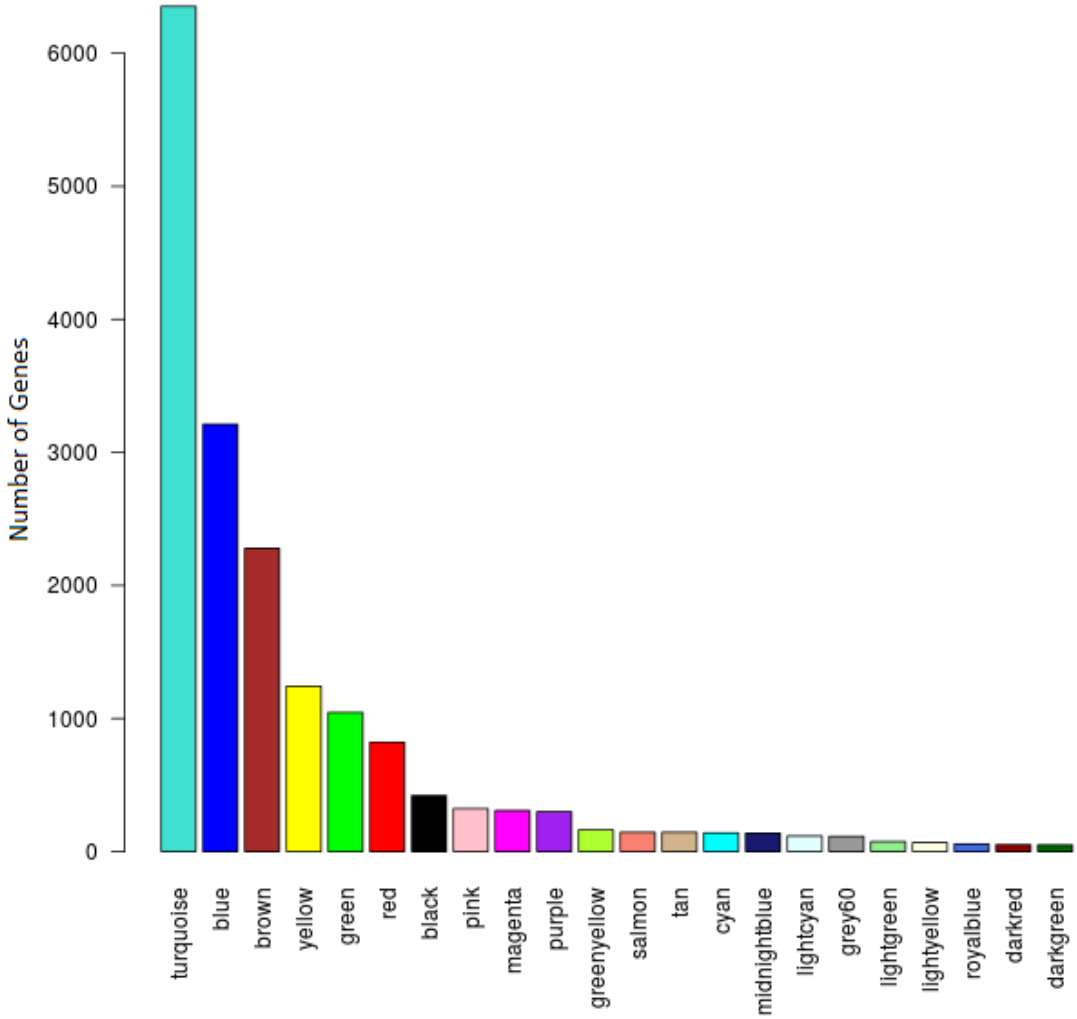


Figure 5.4. Differential co-expression gene modules. Histogram showing cluster sizes (number of genes) per module in the subcortical areas. Gene modules are denoted in different colours.

Differential co-expression gene modules in the cortex.

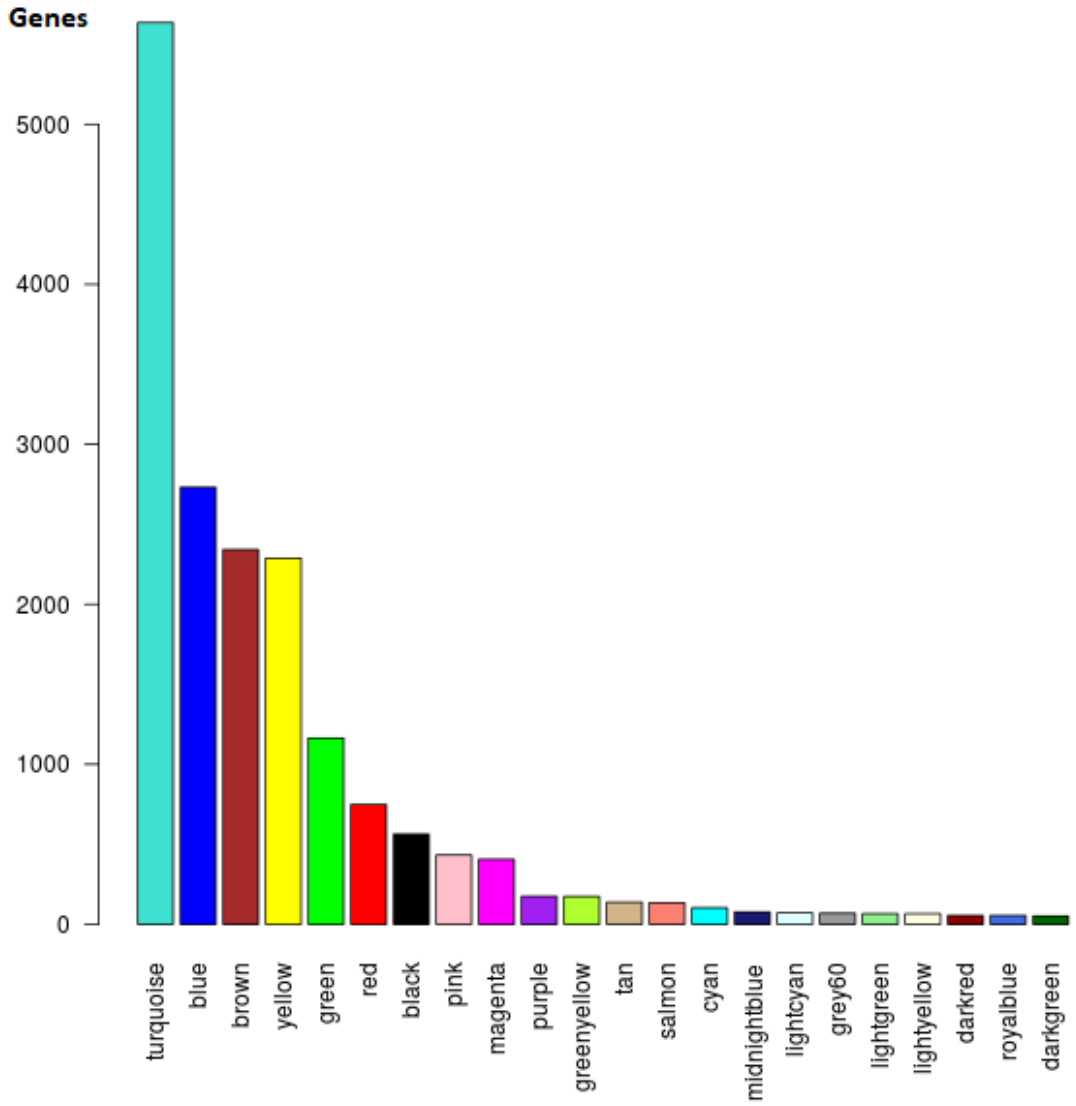


Figure 5.5. Differential co-expression gene modules. Histogram showing cluster sizes (number of genes) per module in the cortex. Gene modules are denoted in different colours.

A close inspection of the correlation heat-maps of the resulting clusters confirms pronounced changes in the correlated structure of each module in the transition between mutant Grb10KO$^{+/p}$ and WT in both subcortical areas (Figure 5.6) and cortex (Figure 5.7). Performing permutation analysis to determine the statistical significance of the observed correlation changes in all modules revealed that the observed changes had been significant for all 22 modules in both the subcortical areas and the cortex.

Running Bonferroni adjusted differential co-expression p-values for each cluster

module in the subcortical areas showed nine significant modules namely: salmon, cyan, light yellow, midnight blue, light green, dark green, royal blue, light cyan and dark red. While in the cortex, 10 modules out of 22 have significant adjusted p-values namely: tan, midnight blue, salmon, light cyan, light green, grey60, dark green, dark red, light yellow and royal blue (Table 5.1). The module where the Grb10 is presented in the subcortical areas (dark red) shows significant Bonferroni adjusted differential co-expression p-value =0, while in the cortex (yellow colour) shows no significant Bonferroni adjusted differential co-expression p-value =1 (Table 5.1).

Heat maps of the Pearson correlation coefficients between all individual gene pairs contained within each module in the subcortical areas.

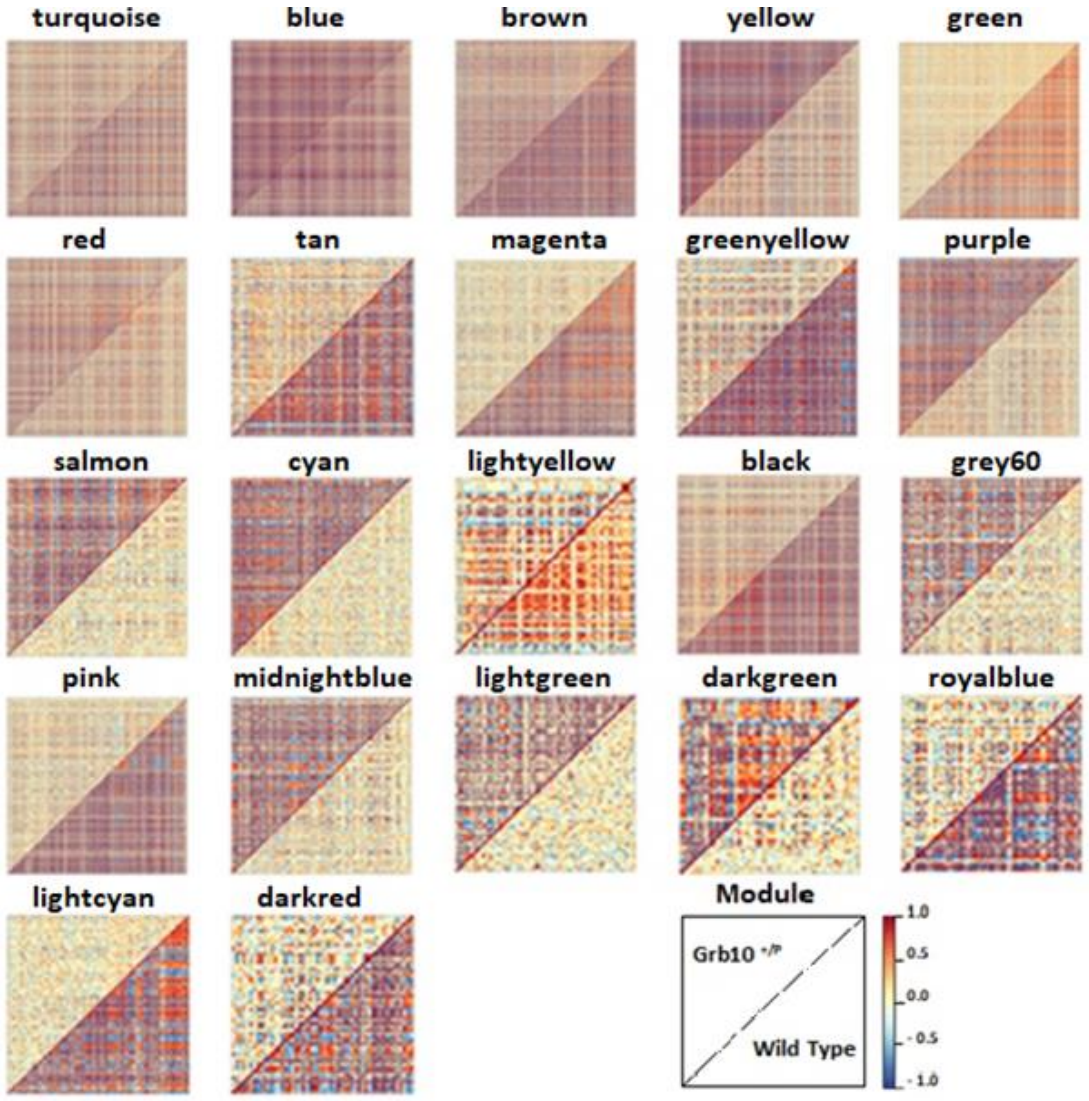


Figure 5.6. Heat maps of the Pearson correlation coefficients between all individual gene pairs contained within each module in the subcortical areas. Each heat map shows mutant Grb10KO$^{+/p}$ and WT co-expression separately (upper and lower diagonal respectively). Colour scale for correlation coefficients is shown at the bottom right corner of this panel.

Heat maps of the Pearson correlation coefficients between all individual gene pairs contained within each module in the cortex.

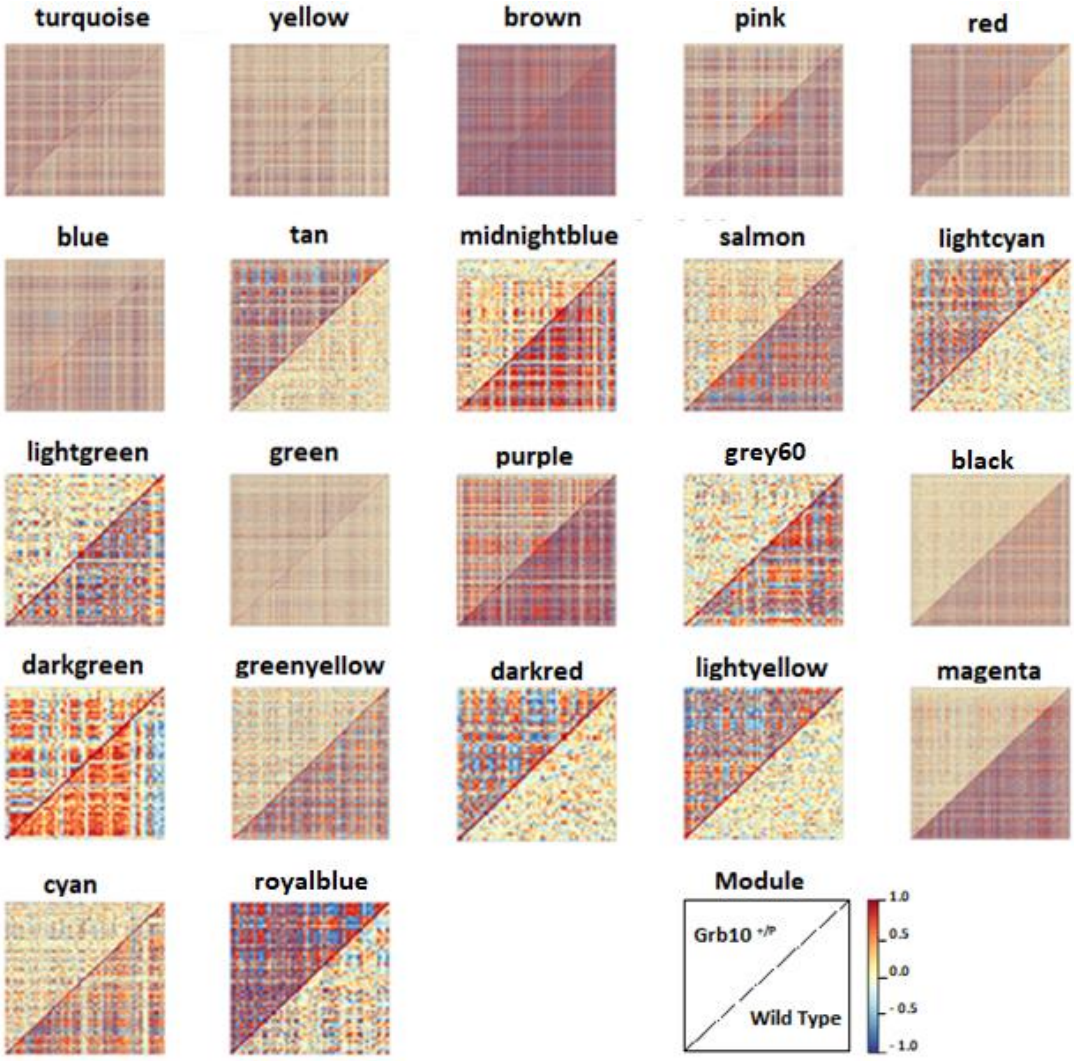


Figure 5.7. Heat maps of the Pearson correlation coefficients between all individual gene pairs contained within each module in the cortex. Each heat map shows mutant Grb10KO$^{+/p}$ and WT co-expression separately (upper and lower diagonal respectively). Colour scale for correlation coefficients is shown at the bottom right corner of this panel.

Distinct cellular functions are enriched in differentially co-expressed clusters

To define the set of biological functions targeted by those modules which have adjusted p-value less than 0.05, the number of GO terms is determined within the biological process category, statistically overrepresented within those modules. The total list of GOIDs for the biological process had revealed n =3550 after selective of the threshold of 10 genes related to the GO term.

Table 5.1. Bonferroni results. Adjusted differential co-expression p-value (Bonferroni) result from lower to higher in each module in the subcortical areas and the cortex.

subcortical areas module	Bonferroni-adj. p	Cortex module	Bonferroni-adj. p
dark green	<0.001	dark green	<0.001
dark red	<0.001	dark red	<0.001
light cyan	<0.001	light cyan	<0.001
light green	<0.001	light green	<0.001
light yellow	<0.001	light yellow	<0.001
midnight blue	<0.001	royal blue	<0.001
salmon	<0.001	salmon	<0.001
cyan	0.022	tan	<0.001
royal blue	0.022	midnight blue	0.022
black	0.066	grey60	0.022
pink	0.066	cyan	0.11
grey60	0.088	green yellow	0.528
yellow	0.154	purple	0.572
Tan	0.22	magenta	0.968
green yellow	0.242	turquoise	1
magenta	0.33	yellow	1
green	0.572	brown	1
purple	0.616	pink	1
Red	0.902	red	1
turquoise	1	blue	1
blue	1	green	1
brown	1	black	1

Gene ontology enrichment analysis in the subcortical areas reveals that the differentially co-expressed modules are enriched in a distinct set of biological processes. Three modules are of particular interest as they show significant enrichment in genes associated with neuronal functions and or development. The salmon colour module in the subcortical areas associated with pituitary gland development, it also plays important role in regulation of cellular protein processes such as regulation of protein localization, establishment of protein localization to plasma membrane, protein heterotetramerization, proteasome binding, positive regulation of cellular protein catabolic process, negative regulation of cyclin-dependent protein serine/threonine kinase activity, ephrin receptor activity, protein

K48-linked deubiquitination (degradation of protein-protein interactions), aggresome & synaptonemal complex assembly. It is as well associated with transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding, RNA polymerase II repressing transcription factor binding, negative regulation of sequence-specific DNA binding transcription factor activity & negative regulation of DNA damage response. Furthermore, this module is associated with embryonic morphogenesis as embryonic skeletal system morphogenesis, embryonic forelimb and hindlimb morphogenesis, hair follicle morphogenesis, cilium morphogenesis & heart morphogenesis (Figure 5.8).

GO enrichment of differentially co-expressed for clustered module (salmon colour) in the subcortical areas.

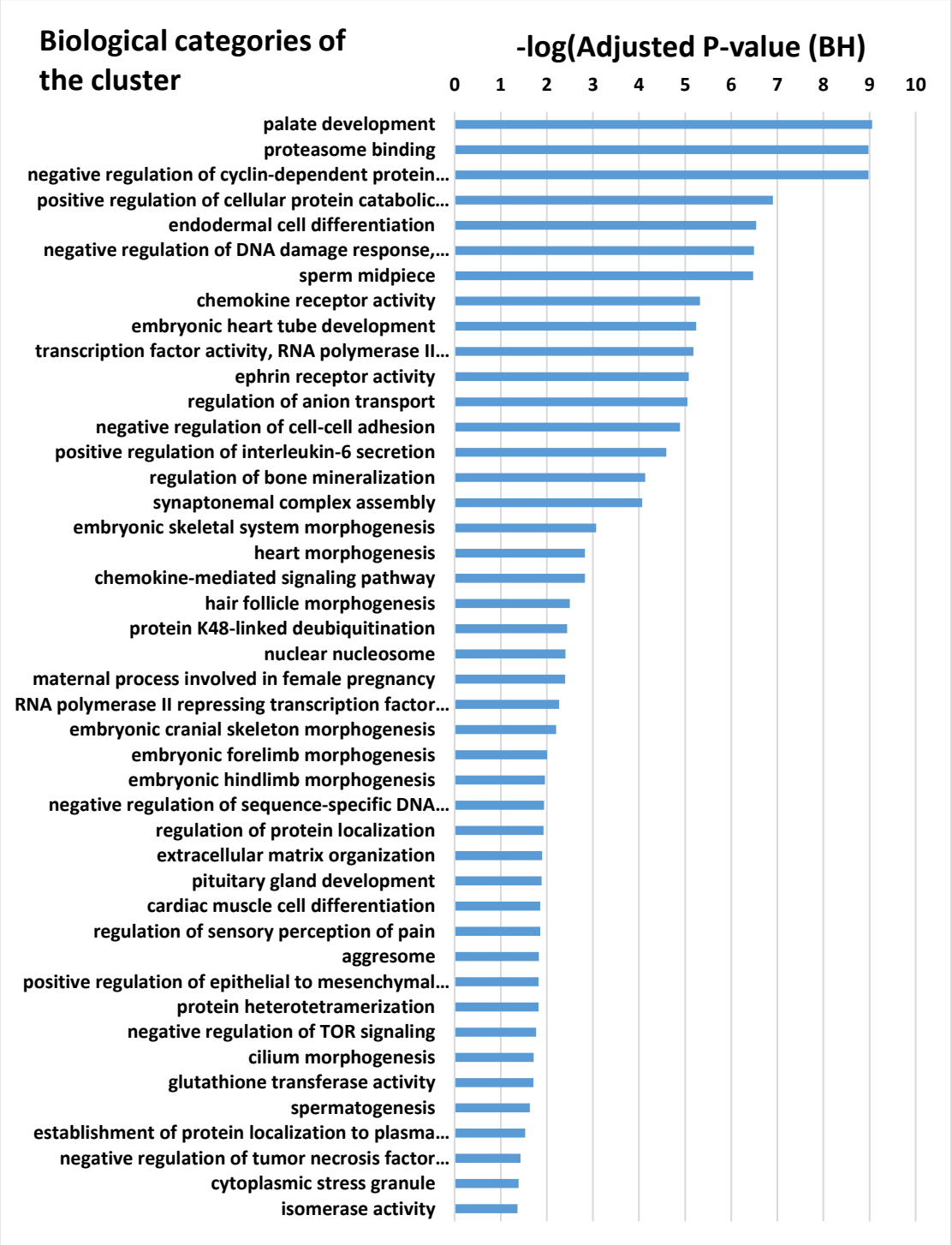


Figure 5.8. Bar plot showing the number of significantly enriched GO terms in the clustered module (salmon colour) in the subcortical areas. Statistical significance in the enrichment of biological process GO terms (=44) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

The cyan colour module in the subcortical areas mostly associated with the developmental functions such as development of epidermis, endoderm and embryonic digestive tract, regulation of cell proliferation, cell migration, apoptosis and peptidyl-tyrosine phosphorylation, receptor tyrosine kinase binding. In the nervous system, it is associated with positive regulation of neuroblast proliferation, negative regulation of interleukin-6 production, oligodendrocyte differentiation, neuron fate commitment. Association with some signalling pathway is also presented such as Notch, extrinsic apoptotic, insulin receptor and protein kinase B signalling as shown in Figure 5.9.

The light yellow colour module in the subcortical areas plays role in CNS as it is associated with neural crest cell development, neurotransmitter transport, neurotransmitter: sodium symporter activity & hindbrain development. Moreover, it is associated with skeletal system development such as ossification, positive regulation of bone mineralization, BMP signalling pathway which is a group of protein signalling molecules that belong to growth factors beta and induce bone formation, positive regulation of osteoblast differentiation & transforming growth factor beta receptor binding. It also mediates intracellular signalling initiated by extracellular stimuli, for instance, SMAD protein signal transduction, positive regulation of pathway-restricted SMAD protein phosphorylation, regulation of MAPK cascade, Wnt-protein, cellular response to an extracellular stimulus, coupled to transmembrane movement of substances, transmembrane transport & transmembrane transporter activity. In addition, this module is associated with transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding. Beside those, this module associated with meiotic cell cycle for example synapsis (fusion of chromosome pairs at the start of meiosis), reciprocal meiotic recombination (separation of chromosomes) & synaptonemal complex assembly (aligned pairs of homologous chromosomes during meiosis) (Figure 5.10).

GO enrichment of differentially co-expressed for clustered module (cyan colour) in the subcortical areas.

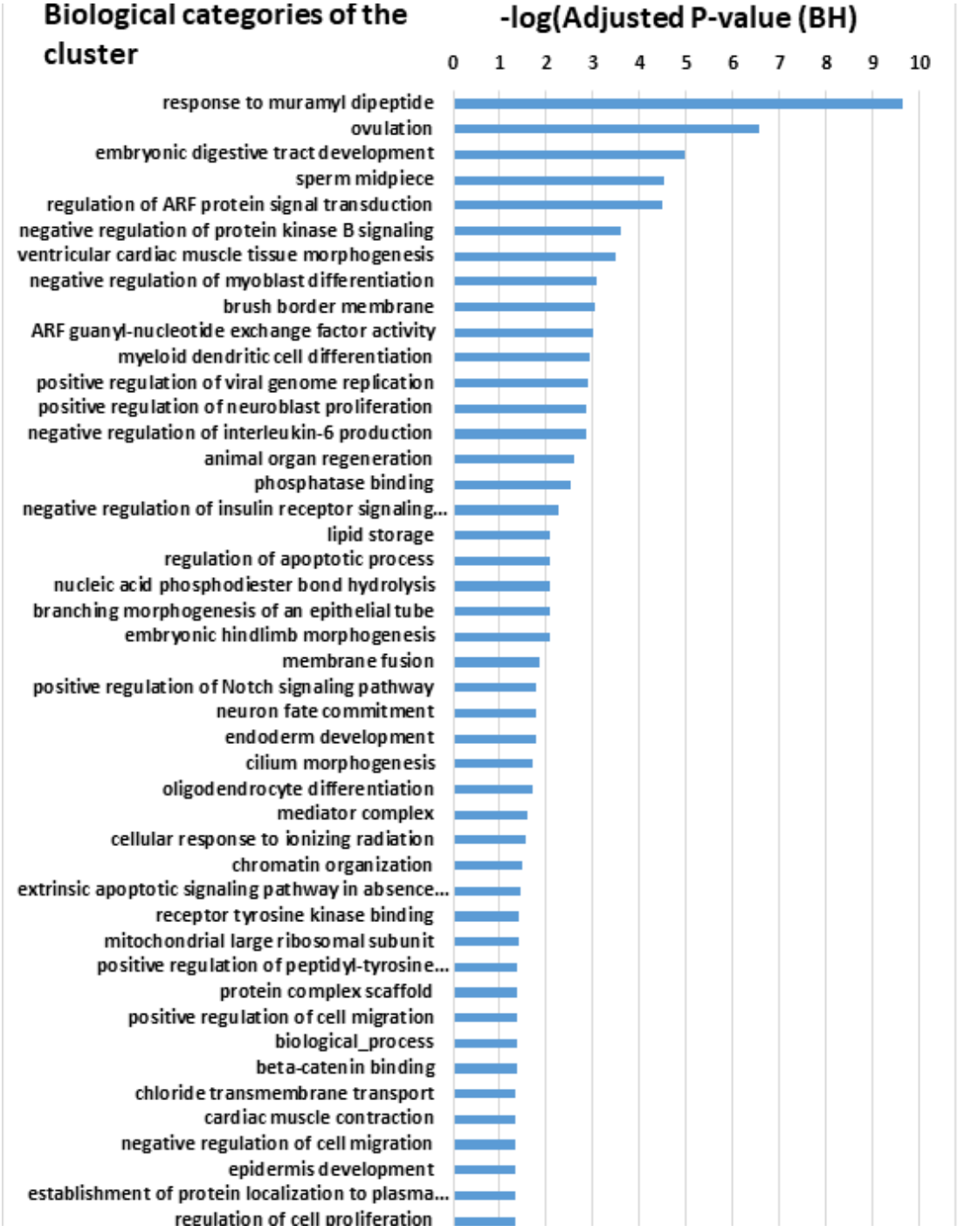


Figure 5.9. Bar plot showing GO enrichment of differential co-expressed. Clustered module (cyan colour) in the subcortical areas. Statistical significance in the enrichment of biological process GO terms (=45) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

Grb10 is part of the dark red colour module in the subcortical areas with a significant difference in co-expression in the KO compared to the WT co-expression profile (Table 5.1). This cluster is significantly enriched in metabolic functions including insulin-like growth factor receptor signalling pathway, G-protein coupled acetylcholine receptor signalling pathway & GTPase activator activity, lipid catabolic process, Rho guanyl-nucleotide exchange factor activity and regulation of Rho protein signal transduction (Figure 5.11).

Examining functional enrichment in modules of differentially expressed genes in the neocortex, three modules are of particular interest. The tan colour module is enriched in associations with neuronal development including, positive regulation of dendritic spine development, positive regulation of axonogenesis & pituitary gland development. In addition, it is associated with the immune response such as positive regulation of isotype switching to IgG isotypes and positive regulation of B cell & T cell differentiation. It also shows a significant association in the Notch signalling pathway (Figure 5.12). The midnight blue colour module in the cortex is significantly enriched in associations also related to neuronal development including synapse organization, positive regulation of neuron apoptotic process. With extracellular region, it shows association with extracellular space, extracellular matrix, extracellular matrix organization, proteinaceous extracellular matrix & cell-matrix adhesion. Besides its association with positive regulation of protein binding, protein processing and MHC class II protein complex (Figure 5.13). The light green module colour in the cortex associated with neurotransmitter transport, beta-amyloid binding, Moreover, the module is associated with cellular protein catabolic process, ubiquitin-dependent protein catabolic process & proteolysis (Figure 5.14).

Grb10 is found in the yellow colour module in the cortex but this cluster is not significantly different when comparing WT and KO matrices which is consistent with the fact that Grb10 is only marginally expressed in the neocortex and therefore we don’t expect a significant shift in co-expressed partners in the KO mice (Figure 5.15).

GO enrichment of differentially co-expressed for clustered module (light yellow colour) in the subcortical areas. *Only top 50 categories are represented.

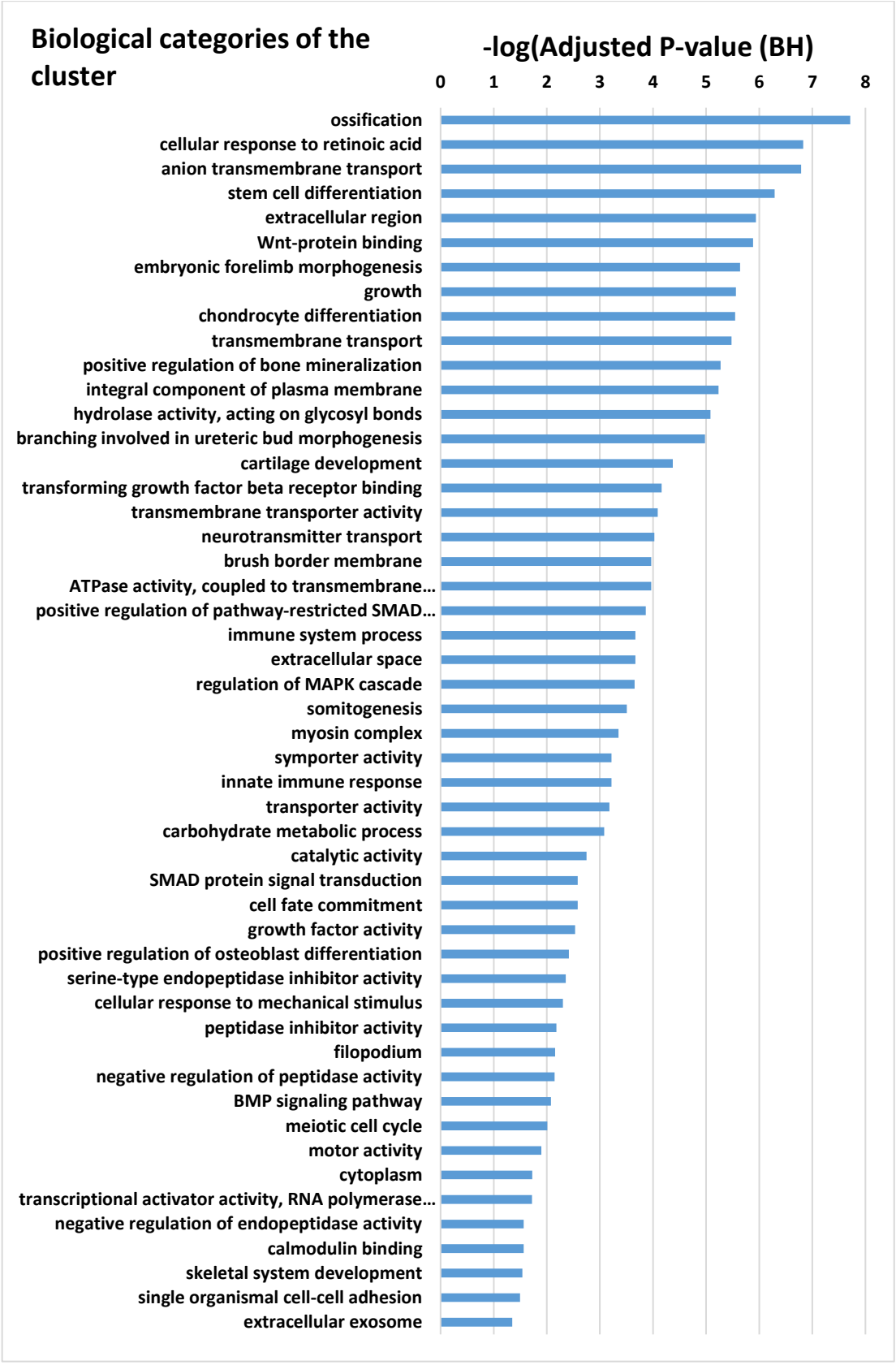


Figure 5.10. Bar plot showing the number of significantly enriched GO terms in the clustered

module (light yellow colour) in the subcortical areas. Statistical significance in the enrichment of biological process GO terms (=67) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

GO enrichment of differentially co-expressed for clustered module (dark red colour) in the subcortical areas where Grb10 is presented.

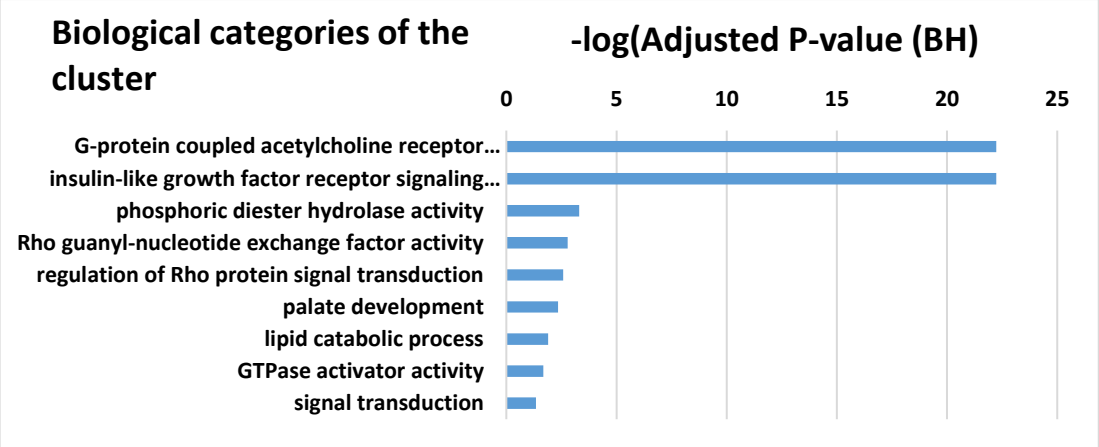


Figure 5.11. Bar plot showing the number of significantly enriched GO terms in the clustered module where Grb10 is presented (dark red colour) in the subcortical areas. Statistical significance in the enrichment of biological process GO terms (=9) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

GO enrichment of differentially co-expressed for clustered module (tan colour) in the cortex.

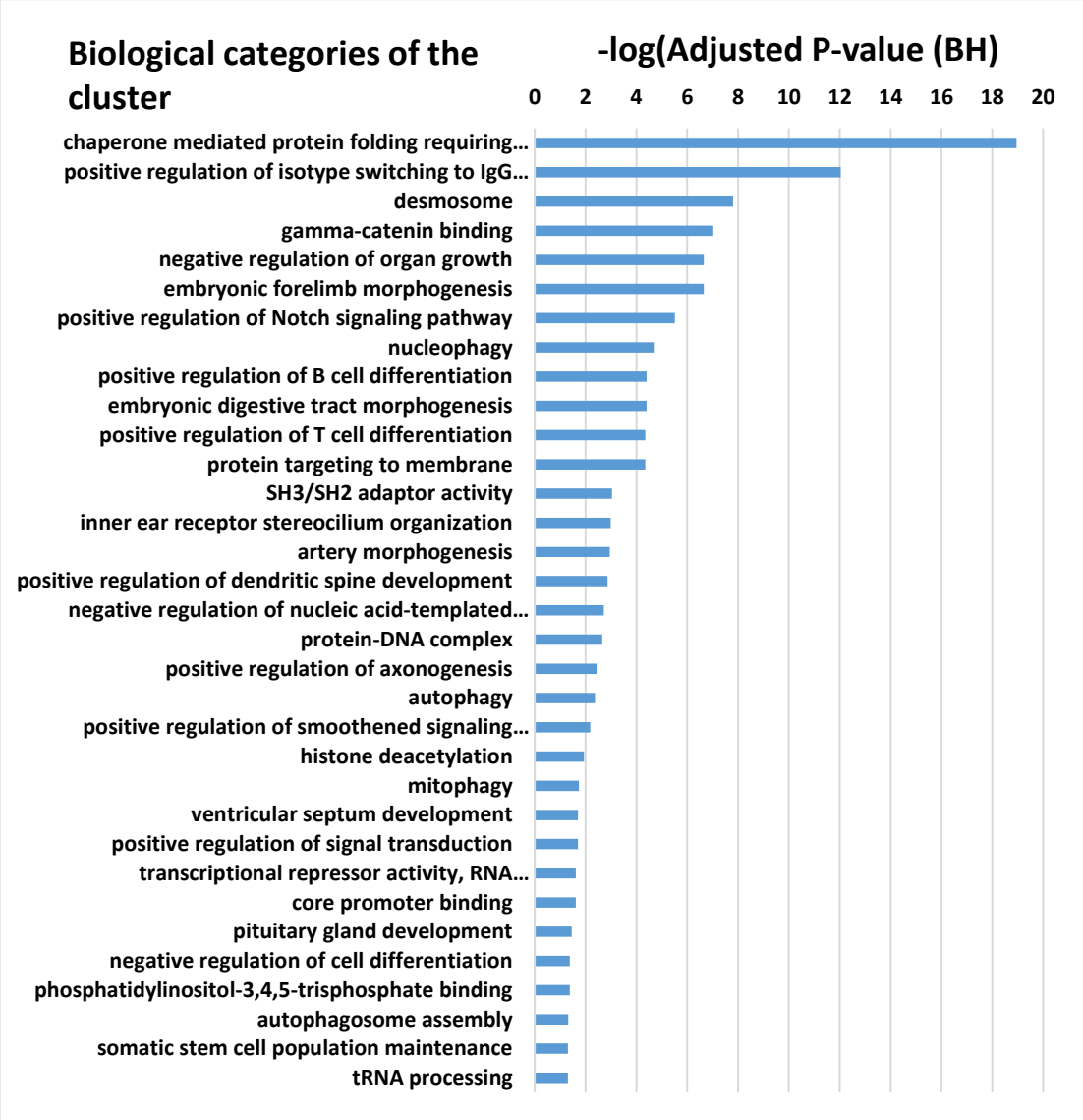


Figure 5.12. Bar plot showing the number of significantly enriched GO terms in the clustered module (tan colour) in the cortex. Statistical significance in the enrichment of biological process GO terms (=33) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

GO enrichment of differentially co-expressed for clustered module (midnight blue) in the cortex.

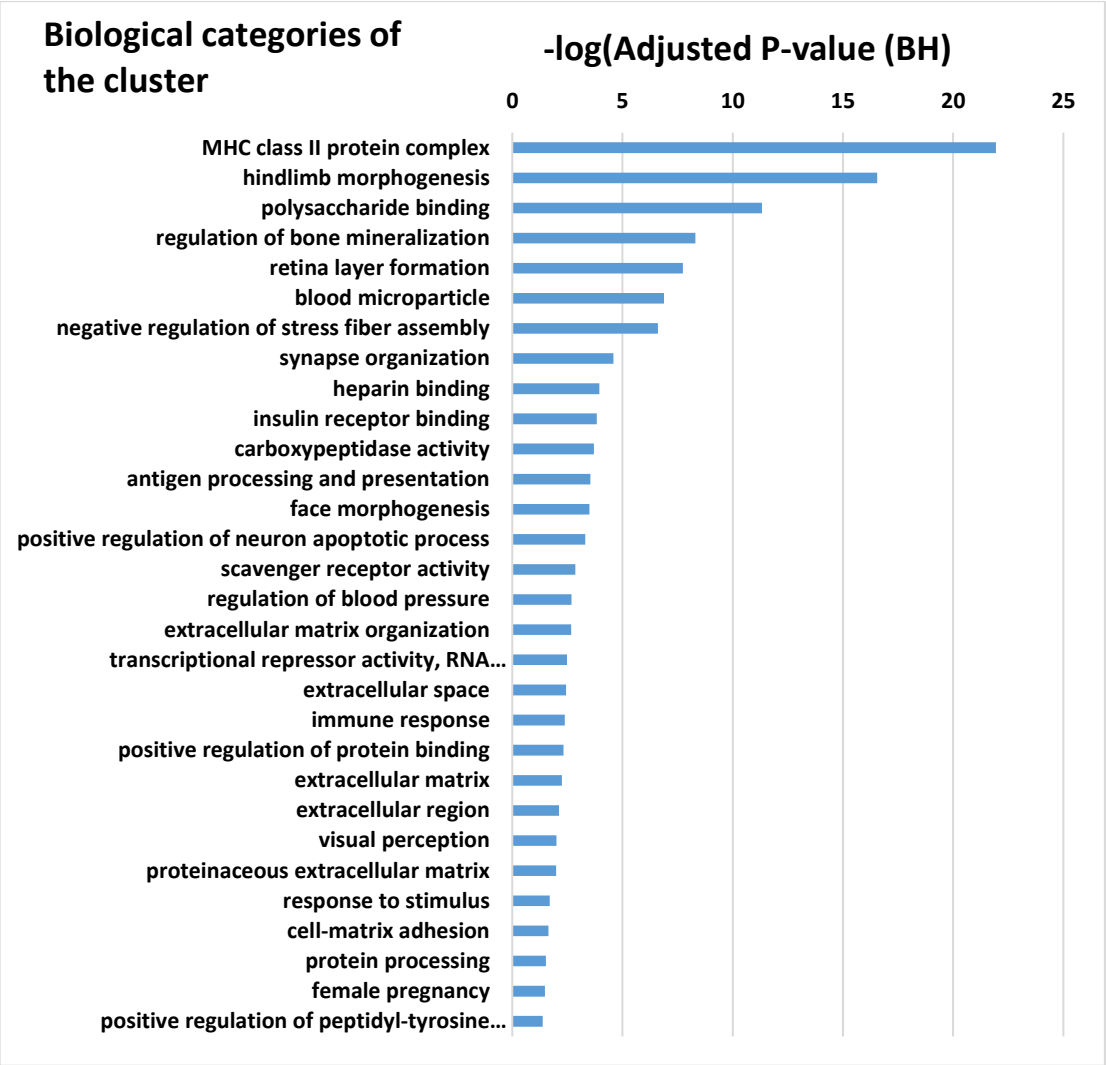


Figure 5.13. Bar plot showing GO enrichment of differential co-expressed (Bonferroni adjustment) clustered module (midnight blue colour) in the cortex. Statistical significance in the enrichment of biological process GO terms (=30) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

GO enrichment of differentially co-expressed for clustered module (light green colour) in the cortex.

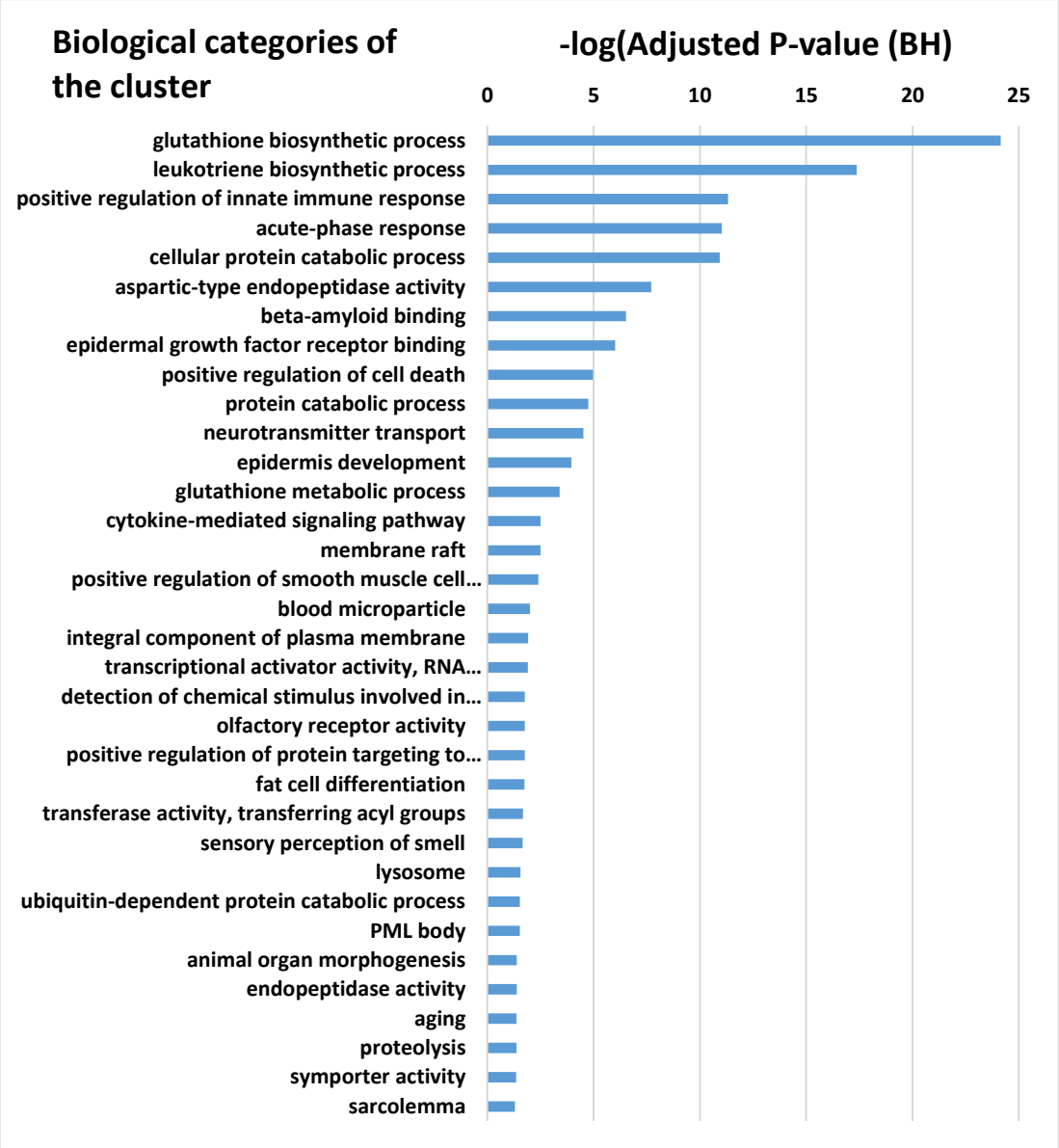


Figure 5.14. Bar plot showing GO enrichment of differential co-expressed (Bonferroni adjustment) clustered module (light green colour) in the cortex. Statistical significance in the enrichment of biological process GO terms (=34) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

GO enrichment of differentially co-expressed for clustered module (yellow colour) in the cortex where Grb10 is presented.

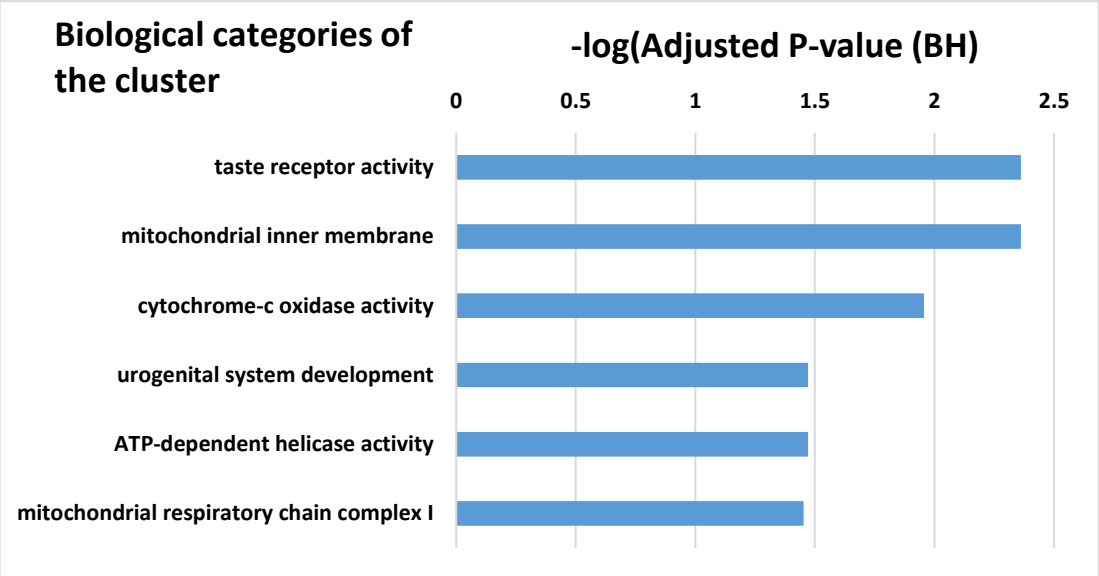


Figure 5.15. Bar plot showing the number of significantly enriched GO terms in the clustered module where Grb10 is presented (yellow colour) in the cortex. Statistical significance in the enrichment of biological process GO terms (=6) was numerically assessed by comparing with 10 000 equally sized random samples of genes. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

Co-expression network of the 30 genes in the subcortical areas.

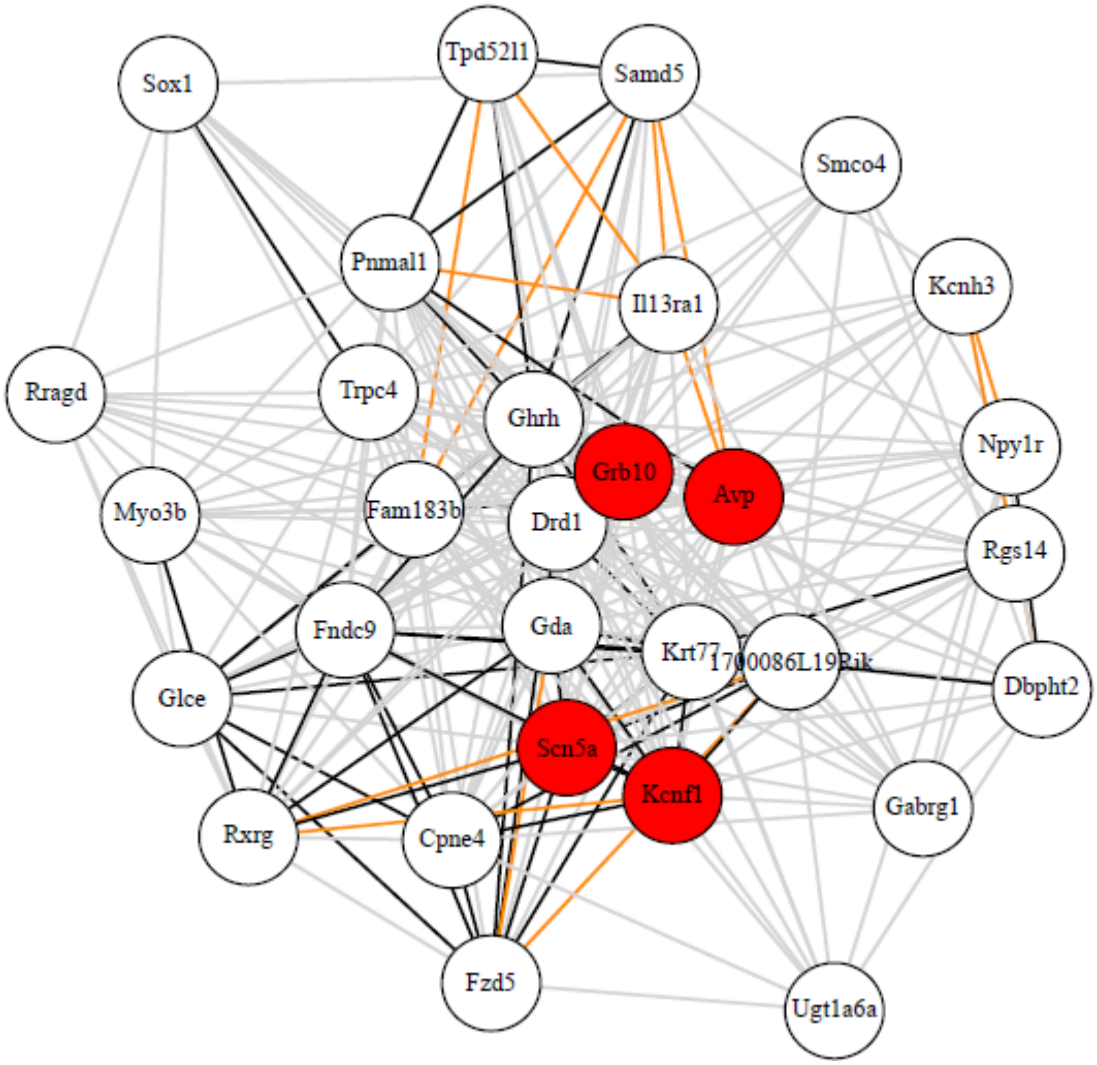


Figure 5.16. Co-expression network of the 30 genes in the subcortical areas. The nodes represent the group of genes most co-expressed with Grb10 in the subcortical areas in WT mice. Red colour nodes represent significantly down-regulated genes and white colour represent genes with no significant changes. The edges represent co-expression between the pairs of genes with a threshold of 0.8 or higher, black edges represent co-expression in WT and mutant conditions, grey edges represent co-expression only in WT mice, and orange edges represent co-expression only in mutant mice.

Co-expression network of the 30 genes in the cortex.

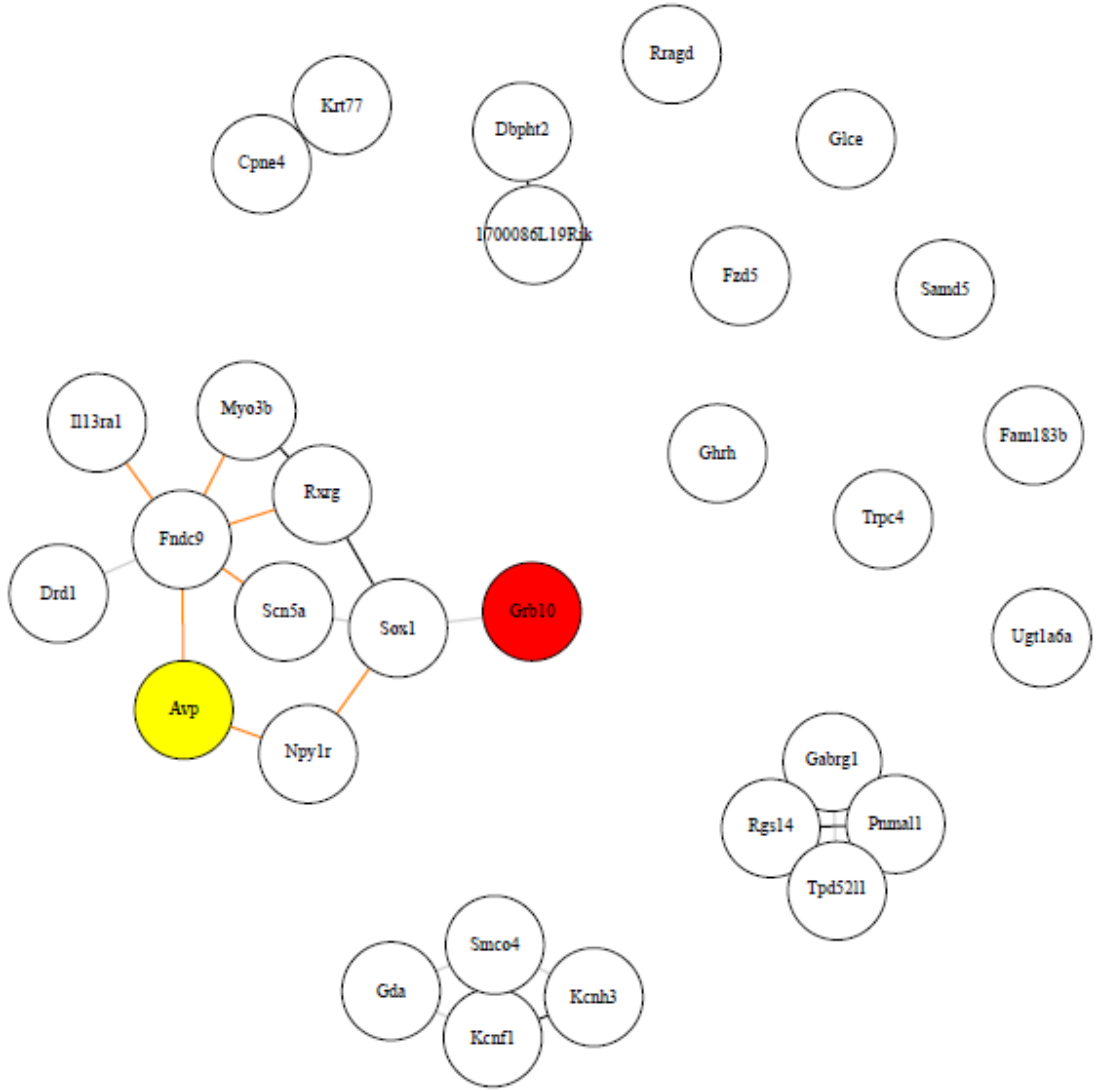


Figure 5.17. Co-expression network of the 30 genes in the cortex. The nodes represent the same group of genes which are most co-expressed with Grb10 in the subcortical areas in WT mice. Red colour nodes represent significantly down-regulated genes, yellow colour represent significantly up-regulated genes and white colour represent genes with no significant changes. The edges represent co-expression between the pairs of genes with a threshold of 0.8 or higher, black edges represent co-expression in WT and mutant conditions, grey edges represent co-expression only in WT mice, and orange edges represent co-expression only in mutant mice.

Table 5.2. GO categories of the 30 co-expressed genes (including Grb10) with correlation value of 0.8 or higher in the WT and its corresponding in the mutant in the subcortical areas and the cortex.

GOID	GO Term	Adj.P-value
GO:0042711	maternal behaviour	1.36E-106
GO:0007625	grooming behaviour	3.84E-56
GO:0030819	positive regulation of cAMP biosynthetic process	3.18E-53
GO:0019228	neuronal action potential	3.95E-44
GO:0007612	learning	7.47E-26
GO:0007631	feeding behaviour	3.78E-25
GO:0007626	locomotor behaviour	3.75E-24
GO:0060291	long-term synaptic potentiation	2.20E-22
GO:0007189	adenylate cyclase-activating G-protein coupled receptor signalling pathway	6.84E-18
GO:0008542	visual learning	4.08E-16
GO:0034220	ion transmembrane transport	1.39E-12
GO:0006811	ion transport	5.10E-08
GO:0006813	potassium ion transport	7.06E-07
GO:0034765	regulation of ion transmembrane transport	1.18E-06
GO:0071805	potassium ion transmembrane transport	1.34E-06
GO:0014070	response to organic cyclic compound	2.78E-06
GO:0055085	transmembrane transport	1.23E-05
GO:0007166	cell surface receptor signalling pathway	0.0013354
GO:0006810	transport	0.00322723
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	0.01044046
GO:0008284	positive regulation of cell proliferation	0.01770826
GO:0007165	signal transduction	0.04444254

Discussion

The development of the nervous system is a highly complex process and gene expression patterns vary widely depending on the specific developmental trajectories of different neural structures (Stead et al., 2006; Sterner et al., 2012). The use of genomics tools has revolutionized the field of behavioural genetics by providing tools to measure the dynamic nature of brain gene expression in relation to behaviour.

Mutant paternal Grb10KO causes social dominance in mice, this might provide a good platform for investigating the relationship between brain gene expression and social behaviour which is a complex phenotype that is developmentally labile. Changes in gene expression profiles during development are entirely determined by an underlying network of precise regulatory interactions between individual genes. I asked about

the type of gene regulatory network and the correlation between sets of genes at different developmental stages in WT and mutated Grb10KO$^{+/p}$ mice in the subcortical areas and the cortex. To address this problem, I focused on the co-expression structure of the transcriptome as a measure of its regulatory network at defined stages of the development (E18.5, 1W, 1M, 3M and 6M). I used existing expression data derived from the mutant Grb10$^{+/p}$ and WT at different developing stages of mice brain mainly the subcortical areas and the cortex and examine the relationship between changes in gene expression profiles and the underlying correlation structure developmental transcriptome.

To distinguish between these two regions of the brain I compared the co-expression structure of the transcriptome between WT and mutant Grb10KO$^{+/p}$. Analysis of gene co-expression has been used extensively to get knowledge about the functional organization of transcriptomes throughout species, tissues and conditions (Yu et al., 2003; Oldham et al., 2006; Oldham et al., 2008; Marco et al., 2009; Saris et al., 2009; Usadel et al., 2009; Torkamani et al., 2010; Kang et al., 2011; Obayashi & Kinoshita, 2011; Zhang et al., 2012A; Zhang et al., 2013; Gaiteri et al., 2014), and the transcriptome of the developmental brain showed distinct co-expression networks which are presumed to act as single expression units consisting of coherent groups of coregulated genes (Kang et al., 2011) and expected to have a constant feature of the normal developmental programme.
Hence, changes in the gene co-expression network have been linked to regulatory dysfunctions which are related to the progression or onset of different pathological and disease conditions such as neurodegeneration and neuropsychiatric disorders, cancer, obesity and possible genome instability associated with age-related functional decline (Miller et al., 2008; Saris et al., 2009; Southworth et al., 2009; Torkamani et al., 2010; Voineagu et al., 2011; de Jong et al., 2012; Ponomarev et al., 2012; Chen et al., 2013; Zhang et al., 2013). This link between pathological dysfunctions and changes in the global network of regulatory interactions suggest that the stability of the co-expression structure of the transcriptome is an essential condition for the normal function of cells and tissues.

Therefore, under the constant network of regulatory interactions, a little difference in the co-expression structure across those two regions of the brain between WT and mutant Grb10KO$^{+/p}$ was predicted. Our result reveals 22 modules in each of the subcortical areas (Figure 5.6) and the cortex (Figure 5.7). Nine of these modules in

the subcortical areas and ten modules in the cortex showed adjusted differential co-expression p-value (Bonferroni) less than 0.05 (Table 5.1) between the WT and mutant $Grb10^{+/p}$. The module where Grb10 is presented showed significant adjusted p-value in the subcortical areas but not in the cortex. This result may be due to the fact that in the cortex Grb10 is less expressed and that's why not supporting the Grb10 as a part of a cluster showing no significant difference in the cortex.

Nevertheless, several modules of co-expressed genes in the cortex were shown to be altered. Gene ontology enrichment analysis conducted to gain insight into the functional coherence of those observed significant modules of clusters and whether if they focused particular biological functions. The result revealed that each module targets a discrete set of biological functions with little functional overlap between modules. this result shows that the transcriptome in the subcortical areas and the cortex are organized into discrete clusters each involved in the regulatory defined sets of biological functions. This transition is organized into separate sets of biological functions suggests new functional associations involved in the normal transition in different developmental stages of the brain. In addition, the fact that the transcripts in some of the significantly differential co-expression modules of the subcortical areas and the cortex are enriched in the brain, suggesting that these genes function mainly in the central nervous system. Nonetheless, the fact that they fall into particular modules indicates that their specific functions differ, or that they are differentially regulated in a temporally or spatially specific manner.

In the modules of both region of the brain, the correlation for growth factor activity had been determined however insulin-like growth factor (IGF) receptor signalling pathway (GO:0048009) has been specified in the dark red module of the subcortical areas. this is confirming that Grb10 is more prominently expressed in the subcortical areas than the cortex where Grb10 plays important role in plays a functional role in IGF (O'Neil et al.1996; Morrione et al., 1997; Dufresne & Smith, 2005; Li et al., 2013; Yang et al., 2016). Finally, the adjusted differential co-expression p-value indicated 5 out of 9 modules in the subcortical areas and 6 out of 10 modules in the cortex showed biological processes related with the CNS suggesting that Grb10 is one of the architect network in the subcortical areas in the CNS.

To understand further Grb10 functionally related genes, we integrated co-expression network between 30 co-expressed genes (including Grb10) that has correlation value

of 0.8 or higher in the WT and its corresponding in the mutant in both brain tissue, thus complementary information to understand Grb10 function network. The result showed completely different layer of regulation in between both brain tissue. In the subcortical areas the correlation is obvious with dynamic network between Grb10 and all those genes (Figure 5.16) while in the cortex Grb10 showed a network among 9 genes only (Figure 5.17). This is maybe because Grb10 is mainly expressed in the subcortical areas and that knocking out of Grb10 affect its expression in the cortex and or it has a very slight effect or expression in the cortex. With these results, we can conclude that the phenotypic variability associated with social dominant in these genotypes is mostly determined by genotypic variability.

Enrichment analysis for those 30 correlated genes (Table 5.2) reveals their association with behaviours such as maternal behaviour, grooming behaviour, feeding and others. they also associated with signal transduction, ion transport such as potassium and with CNS such as long-term synaptic potentiation and neuronal action potential. They are also associated with growth such as positive regulation of cell proliferation as the growth factors play important roles in regulating cell proliferation (Rozengurt, 1992; Bhora et al., 1995; Li & Lu, 2005).
This association indicates the main function of Grb10 in controlling the growth and social dominant that showed in co-expression network in the subcortical areas (Figure 5.16) where it is mainly expressed.

Chapter VI Comparing maternal and paternal Grb10KO

Introduction

Grb10 (growth factor receptor-bound protein 10) is an imprinted gene which encodes an intracellular signalling adaptor protein. When expressed from the maternal allele, Grb10 acts to restrict foetal growth and is permissive for adipose deposition in adulthood as maternal Grb10 knockout mice exhibit overgrowth in numerous organs and tissues (Charalambous et al., 2003), whilst the paternal allele expressed Grb10 in neurons mainly the subcortical areas and it plays an important role in social dominance behaviour in mice in adulthood (Garfield et al., 2011).

In mice, Grb10 gene expression is imprinted with maternal expression observed in all tissues except the brain. Grb10 influences foetal and placental growth, where the maternal Grb10 allele acts as an inhibitor of both foetal and placental growth (Charalambous et al., 2003; Garfield et al., 2011; Cowley et al., 2014). Disruption of Grb10 gene expression in peripheral tissues leads to significant overgrowth of the mice, indicating a role for endogenous Grb10 as a growth suppressor (Wang et al., 2007). In adult WT mice, Grb10 expression could be detected in tissues, including brain, heart, pancreas, fat, muscle, lung, and spleen (Wang et al., 2007) while in the liver, Grb10 could be detected in embryonic mouse stage (Charalambous et al., 2003) where $Grb10^{m/+}$ mice showed persistent overgrowth of the liver during embryonic stage (Charalambous et al., 2003; Garfield et al., 2011; Cowley et al., 2014) and at birth (Charalambous et al., 2003; Garfield et al., 2011) but has no (Gruppuso et al., 2000) or little (Wang et al., 2007) effect in adult stage. $Grb10^{m/+}$ mice had no effect on Grb10 expression in the brain (Wang et al., 2007) with the paternal allele found to be expressed in subcortical areas from embryonic stages up to adulthood (Garfield et al., 2011). Maternal Grb10 expression in the brain has been found only marginally.

The molecular interactors of Grb10 have been better characterised for the maternal knockout. Grb10 is an adaptor protein that inhibits tyrosine phosphorylation of the insulin, EGF (Ooi et al., 1995), growth hormone (Moutoussamy, 1998) and other growth factor receptors (Morrione, 2003; Nantel et al., 1999; Kabir and Kazi, 2014). It has role in apoptosis signalling (Hu et al., 2010), and has an effect on cell survival (Holt & Siddle, 2005) by binding with cell molecules affecting cell survival such as the IGF-1 receptor (Vecchione et al., 2003; Holt & Siddle, 2005; Cao et al., 2008; Des buquois et al., 2013), RAF-1 (Nantel et al., 1998; Nantel et al., 1999; Kebache et al., 2007), MEK1 (Nantel et al., 1998; Nantel et al., 1999), AKT kinases (Jahn et al., 2002;

Wick et al., 2003; Urschel et al., 2005), the 14-3-3 adaptor protein (Urschel et al., 2005) and NEDD4 ubiquitin ligase (Vecchione et al., 2003; Cao et al., 2008).

In the previous chapters, I have obtained and analysed brain transcriptome data for the paternal Grb10 knockout. The results obtained revealed a series of potential interactors of the paternal Grb10 allele. In this chapter, I compare these results to those obtained in liver tissue comparing wild-type and maternal Grb10KO $^{m/+}$ in order to gain insights into the similarities and differences in the molecular interactions of both alleles.

Aim

The main aim of this chapter is to compare patterns of differential gene expression of Grb10KO of the paternal copy in the brain with the changes observed associated with the maternal KO transcriptome profiles in the liver. For this, I annotated liver transcriptomes of Grb10KO$^{m/+}$ and wild-type mice at E18.5 developmental stage, carried out differential gene expression analyses of maternal mutant by comparing the wild-type and the mutant mice Grb10KO$^{m/+}$ and finally, contrast these results to the paternal mutant Grb10KO$^{+/p}$ mice.

Methods

Sample collection and genotyping

To identify potential gene interactors of maternal Grb10. Mice liver samples at age of E18.5. were collected. In this chapter, I describe the methods used for collecting the samples, RNA preparation, RNA sequencing and bioinformatic analyses for annotation, differential expression and GO to contrast these results with the results found in paternal mutant Grb10KO$^{+/p}$ mice. For this study, three maternal mutant Grb10KO$^{m/+}$ and 3 WT male mice were used. The knockout line used in these mice is that reported by Garfield et al. (2011). Mice were dissected and the liver organ was immediately kept in a separate Eppendorf tube filled with RNAlater to maintain RNA integrity during storage of the samples. Please refer to chapter II, for protocol used

for genotyping.

RNA extraction

For RNA extraction Nucleospin RNA kit (Macherey-Nagel) extraction method was used according to the manufacturer's protocol.

RNA Quality Control

For RNA quality control (purity and concentration) the following procedures were used. First, using a NanoDrop™ spectrophotometer to check A260/280 absorption values, where lower values indicate contamination. Then the samples ran on an Agarose Gel Electrophoresis to check the intact 18S and 28S bands of RNA that can be clearly seen. The samples then purified using the Nucleospin RNA Clean-Up XS protocol and finally stored at -80 °C.

Library preparation and RNA Sequencing

Samples that were selected were sent to collaborators: Kings College London School of Medicine BRC Genomics Core Facility (Guy's and St Thomas' NHS Foundation Trust). Department of Medical and Molecular Genetics 7th Floor, Tower Wing, Guy's Hospital London SE1 9RT. For library preparation and sequencing transcriptome (paired-end reads) Illumina-HiSeq 2500 platform was performed.

Transcriptome profile analyses of the Grb10 maternal $KO^{m/+}$

Quality analysis using FASTQC, transcriptome annotation, analysis of transcriptome splice variants, differential expression analysis, pairwise comparison, and functional enrichment analysis all these methods have been carried out as described in Chapters III and IV.

Results

Transcriptome annotation of Grb10 maternal $KO^{m/+}$ and wild-type controls

The FastQC application of the Galaxy version 0.67 had shown all the paired-end reads samples have a sequence length of 101 bp with zero poor quality sequences. The quality scores in Phred scale for all samples were all above 23 (more than 99% base accuracy). All the sequencing adapters if found had been clipped using a *Cutadapt* tool from the Python packaging index. Kallisto program had been used for annotation of the transcripts. Output file of abundances.tsv was used for further analysis. A total number of 111462 transcripts and 35274 genes were obtained (Table 6.1). Bash command line was used to detect the total number of reads ranged from 61717238 to 67271182 (Table 6.1). Splice variants of Grb10 transcript in the liver showed a total of 240 of three types which are mGrb10-201, mGrb10-203 and mGrb10-206 (Figure 6.1). They are highest and different in comparison with Grb10 transcript in the subcortical areas (mGrb10-208) and the cortex (none) at the stage of E18.5 (Table 6.2).

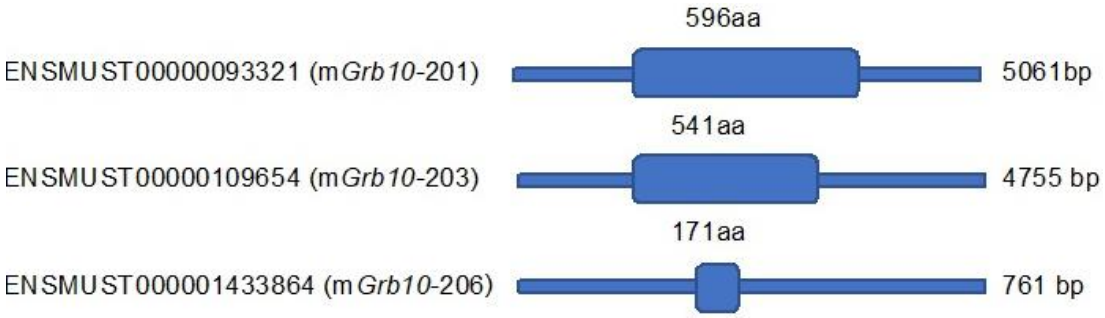


Figure 6.1. Detected Grb10 transcript splice variants. A schematic of the 3 transcript splice variants that have been detected in the liver.

Table 6.1. Annotation. A total number of annotated genes and transcripts using Kallisto and the total number of reads (paired-end reads) in the liver of WT and mutant E18.5 mice (=3 for each).

Total no. of Transcripts	=111462	Total no. of Genes	=35274
Total no. of reads (Forward & Reverse) in WT E18.5 replica (=3) in the liver.	n1 =67271182 n2 =61717238 n3 =62844560	Total no. of reads (Forward & Reverse) in Mutant E18.5 replica (=3) in the liver.	n1 =65119546 n2 =65983318 n3 =62132034

Table 6.2. The edgeR result of transcripts using Kallisto's output. A total number of differentially expressed transcripts (FDR < 0.05), and detected Grb10 splice variants in the liver at E18.5 and its comparison with the subcortical areas and the cortex obtained by Kallisto, ENSMUST00000143386 (mGrb10-206**),** ENSMUST00000148254 (mGrb10-208), ENSMUST00000093321 (mGrb10-201), ENSMUST00000109654 (mGrb10-203).

Different type of Tissue	Subcortical areas	liver	Cortex
FDR_total nos. of transcript at E18.5	214	240	146
Grb10 transcript type	mGrb10-208	mGrb10-201, mGrb10-203 & mGrb10-206	

Maternal Grb10KO$^{m/+}$ leads to differential expression of over 100 genes

The differential expression analysis was carried out using the edgeR package supported in *R* software (Robinson et al., 2010). The analysis was performed on Kallisto's output and the number of down and up-regulated differentially expressed transcriptome and genes obtained (FDR < 0.05) were quantified (Figure 6.2). As expected, Grb10 was detected as differentially expressed in the samples and showed

the highest FDR value (=7.18198335462268E-28) among the down-regulated genes. Overall, the total number of differentially expressed genes (FDR =0.05) is 279 in which 146 genes are up-regulated and 133 genes are down-regulated while the total number of differentially expressed transcripts (FDR =0.05) is 240 (Table 6.2) in which 132 transcripts are up-regulated and 108 transcripts are down-regulated (Figure 6.2).

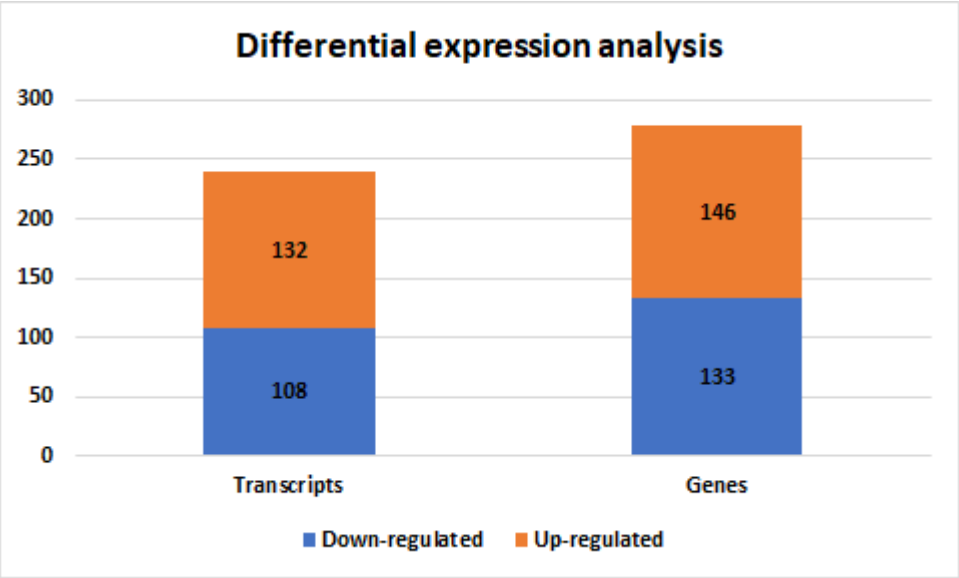


Figure 6.2. Bar chart for the number of differentially expressed genes/transcripts in the liver of E18.5 mice. The bar chart shows the actual size of up- and down-regulated differentially expressed genes/transcripts at E18.5 represented in the liver.

Maternal and paternal KO show distinct transcriptional signatures

In order to assess whether the transcriptional signatures associated with the Knocking out of the paternal allele in the brain tissue are specific to the tissue context or instead a systemic consequence of the decreased Grb10 expression sets of differentially expressed genes in the brain and liver samples were compared. The pairwise comparisons of samples corresponding to E18.5 stage showed two predicted genes (Gm10800 & Gm10722) which were down-regulated in both the liver and the subcortical areas (Figure 6.3 and Table 6.3). No common genes were found to be differentially expressed in the liver and the cortex (Figure 6.3) which is expected as Grb10 paternal copy is only expressed in the subcortical areas with differential gene expression in the cortex for the paternal Grb10KO$^{+/p}$ likely to be a byproduct of the effects of a dearth of Grb10 in the subcortical areas.

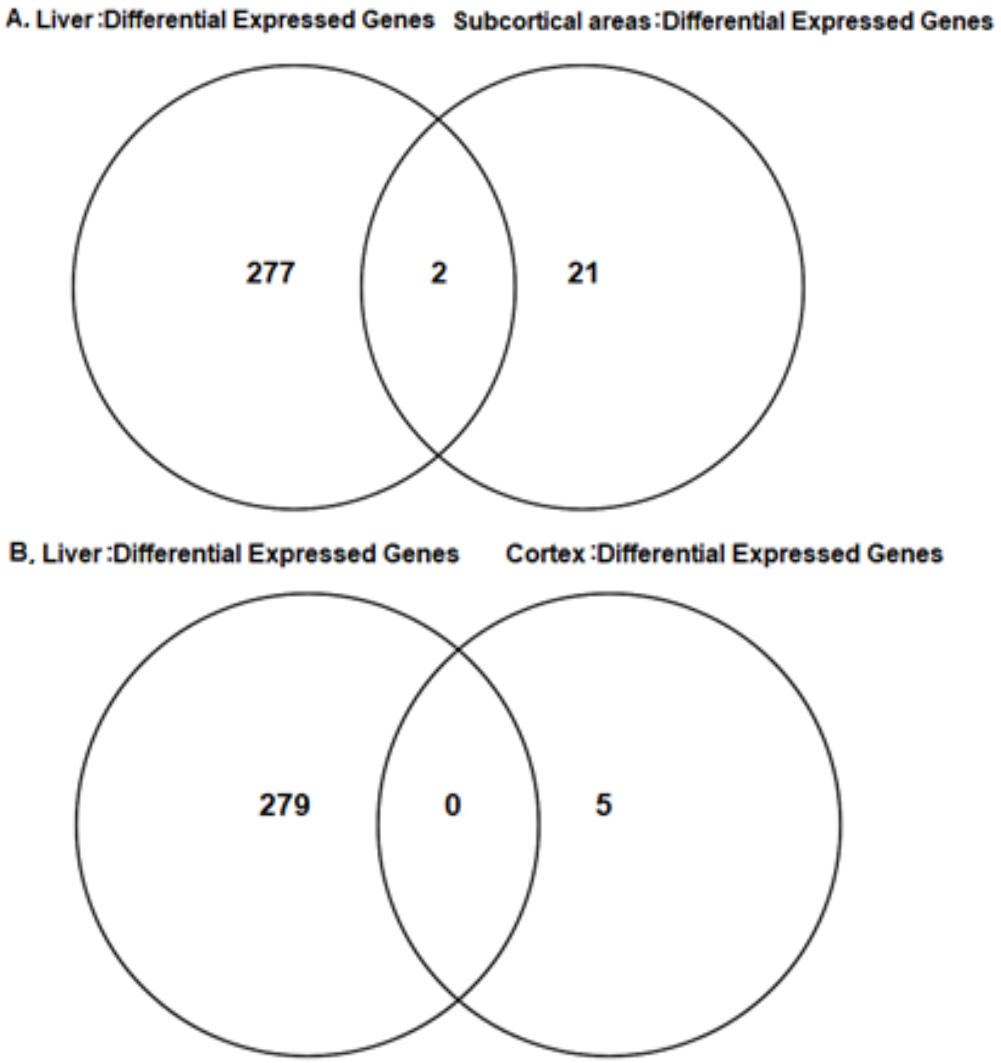


Figure 6.3. Venn diagram summarising the overlap of differentially expressed genes in the liver with a comparison to the subcortical areas (A) and the cortex (B). Pairwise comparison of differentially expressed genes found between the liver & the subcortical areas (A) and between the liver & the cortex (B) in E18.5.

Comparing the differentially expressed genes in E18.5 stage in the liver with the pool of differentially expressed genes in the subcortical areas of the brain across the five developmental stages showed a total of 10 shared differentially expressed genes (Grb10, Slc47a1, Rgs16, Lhx6, Itih2, Ramp3, Gm10800, Gm10722, Gm3375, and Gm15920) in both tissues and 4 shared differentially expressed genes (Grb10, Gm6505, Gm10721, and Gm15920) when comparing the liver and neocortex in which Grb10 and most of the shared genes are down-regulated in all the tissue (Figure 6.4 and Table 6.3).

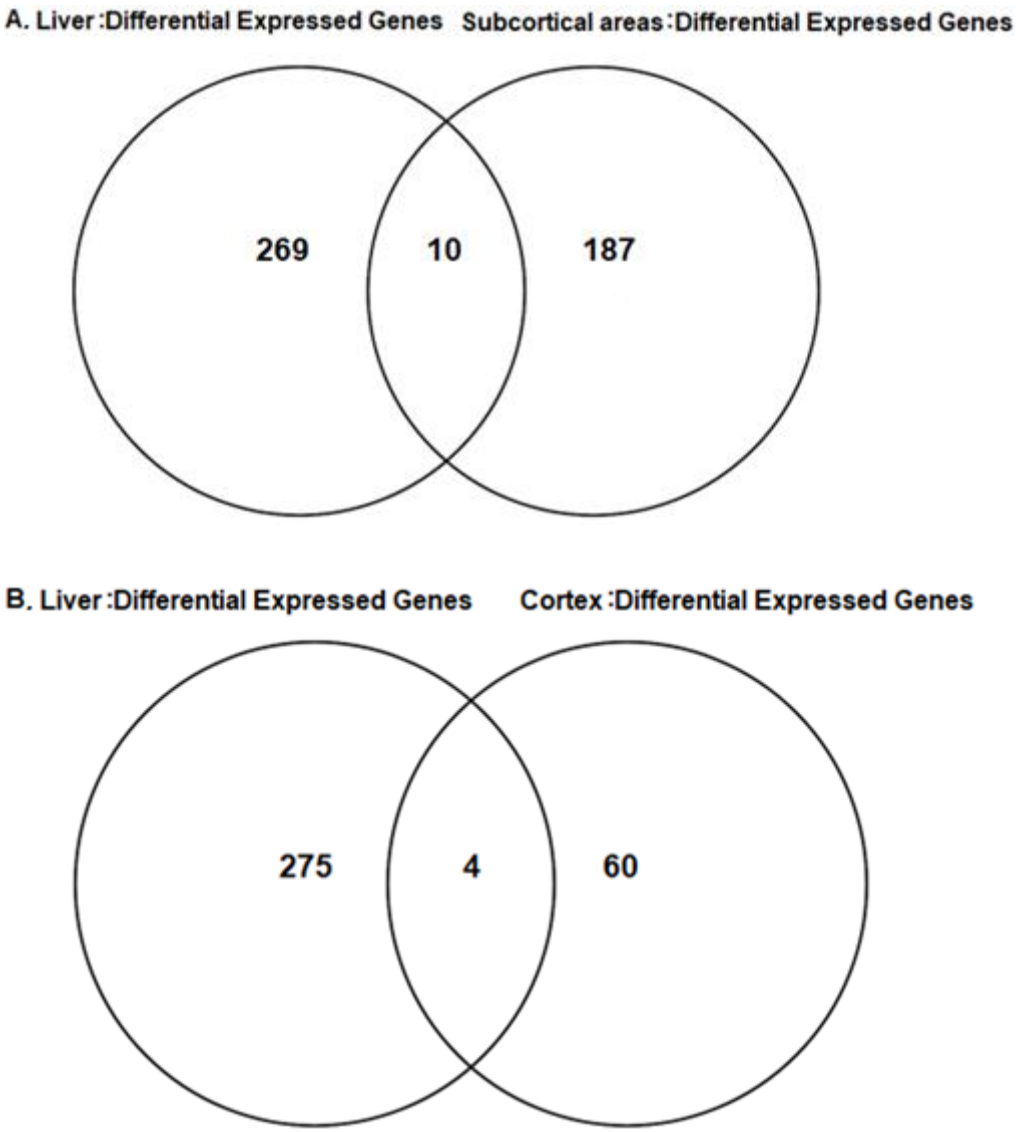


Figure 6.4. Venn diagram summarising the overlap of differentially expressed genes in the liver with a comparison to all the developmental stages in the subcortical areas (A) and the cortex (B). Pairwise comparison of differentially expressed genes found between the liver at E18.5 & all the developmental stages in the subcortical areas (A) and between the liver at E18.5 & all the developmental stages in the cortex (B). The duplicated genes within all the developmental stages were removed.

Table 6.3. Direction of pairwise differentially expressed genes between the liver and brain tissues. The direction of overlapped differentially expressed genes (down- or up-regulated) between the liver and subcortical areas at E18.5, and between the liver at E18.5 and all the developmental stages in the subcortical areas and the cortex.

Tissue/Gene name	Down- or Up-regulated differentially expressed genes	
Liver and Subcortical areas at E18.5		
1- Gm10800 (predicted gene 10800)	Down-regulated	Down-regulated
2- Gm10722 (predicted gene 10722)	Down-regulated	Down-regulated
Liver at E18.5 and Subcortical areas of all stages	**Liver at E18.5**	**Subcortical areas at all stages**
1- Grb10 (Growth factor receptor-bound protein 10)	Down-regulated	Down-regulated
2- Slc47a1 (Multidrug and toxin extrusion protein 1)	Down-regulated	Down-regulated
3- Rgs16 (regulator of G-protein signalling 16)	Down-regulated	Down-regulated
4- Lhx6 (Mus musculus LIM homeobox protein 6)	Down-regulated	Down-regulated
5- Itih2 (inter-alpha-trypsin inhibitor, heavy chain 2)	Down-regulated	Down-regulated
6- Ramp3 (Receptor activity-modifying protein 3)	Up-regulated	Down-regulated
7- Gm10800 (predicted gene 10800)	Down-regulated	Down-regulated
8- Gm10722 (predicted gene 10722)	Down-regulated	Down-regulated
9- Gm3375 (predicted gene 3375)	Down-regulated	Down-regulated
10- Gm15920 (predicted gene 15920)	Up-regulated	Up-regulated
Liver at E18.5 and Cortex of all stages	**Liver at E18.5**	**Cortex at all stages**
1- Grb10 (Growth factor receptor-bound protein 10)	Down-regulated	Down-regulated
2- Gm6505 (predicted gene 6505)	Up-regulated	Down-regulated
3- Gm10721 (predicted gene 10721)	Down-regulated	Up-regulated
4- Gm15920 (predicted gene 15920)	Down-regulated	Up-regulated

Maternal and paternal Grb10 associated with distinct functional categories

The GO enrichment analysis for biological processes category targeted to E18.5 liver which has Bonferroni adjusted p-value less than 0.05 revealed a total of 175 lists of GOIDs, 73 of them are related to down-regulated differentially expressed genes while 102 are related to up-regulated differentially expressed genes.

Gene ontology enrichment analysis reveals that down-regulated differentially expressed genes in the liver targeted the biological processes mostly associated

excretion, digestion, tissue regeneration, cell differentiation, development of organs skeletal muscle, camera-type eye & kidney, some of the metabolic processes such as fatty acid, lipid, lipoprotein, cellular amino acid, triglyceride, steroid (Figure 6.5 and supplementary Figure S6.1). They are associated with biosynthesis processes, metabolic processes, catabolic processes. In addition, genes are associated with response to nutrient, regulation of growth and cell maturation (Figure 6.5 and supplementary Figure S6.1).

The biological process for up-regulated differentially expressed genes in the liver showed GO categories associated with transport such as lipid transport, sodium ion, calcium ion transmembrane transport, inorganic anion transport, organic acid transmembrane transport (Figure 6.6 and supplementary Figure S6.2). They are also associated with some neuronal categories, for example, neuron differentiation, axonogenesis, dendrite development, myelination, regulation of postsynaptic membrane potential and negative regulation of neuron apoptotic process (Figure 6.6). Furthermore, it has shown to be associated with immune response such as neutrophil chemotaxis, response to molecule of bacterial origin, positive regulation of T cell migration. In addition, with some signalling and pathways such as epidermal growth factor receptor signalling pathway, protein kinase B signalling and chemokine-mediated signalling pathway (Figure 6.6 and supplementary Figure S6.2).

There was no significant (Adjusted P-value < 0.05) shared GO term has been found between the liver and each of the brain tissue (the subcortical areas and the cortex) at age of E18.5. However, number of GO terms have been detected among down-regulated differentially expressed genes between the liver and the subcortical areas of all the developmental stages n =14 (Table 6.4). Moreover, only one overlap GO term has been detected among up-regulated differentially expressed genes between the liver and the subcortical areas of all the developmental stages (Table 6.4). In contrast, up-regulated differentially expressed genes in the cortex of all the developmental stages has been shown three shared GO terms with the liver and none of the GO term has been detected with down-regulated differentially expressed genes between the cortex of all the developmental stages and the liver (Table 6.4). All the shared GO term with GOIDs are listed in Table 6.5.

Significantly enriched GO terms of biological process for down-regulated differentially expressed genes in the liver. *Only top 50 categories are represented.

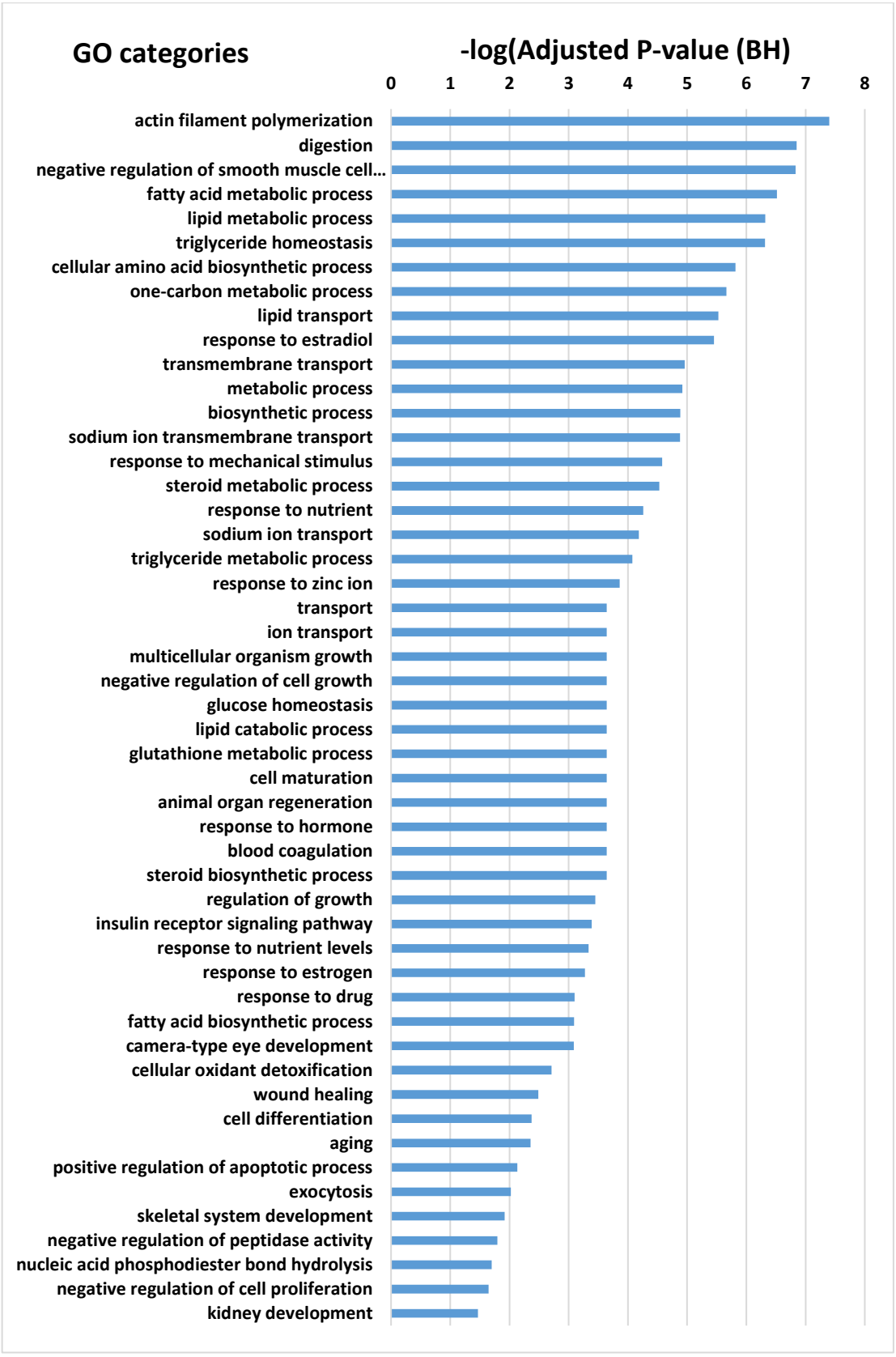


Figure 6.5. Bar plot showing the number of significantly enriched GO terms. Statistical

significance in the enrichment of biological process GO terms (=73) was numerically assessed by comparing with 10 000 equally sized random samples of genes for down-regulated differentially expressed genes in the liver. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

Significantly enriched GO terms of biological process for up-regulated differentially expressed genes in the liver. *Only top 50 categories are represented.

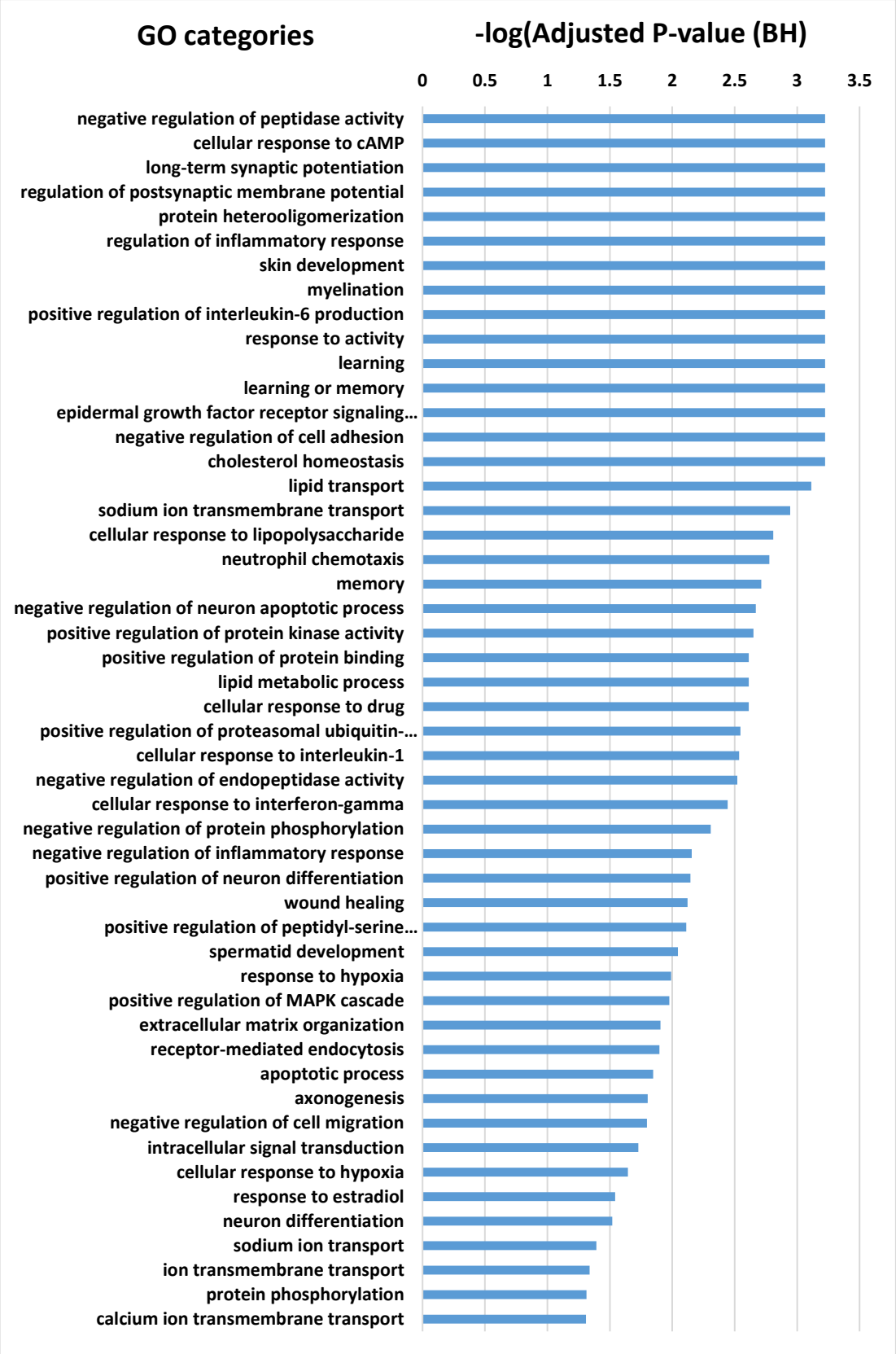


Figure 6.6. Bar plot showing the number of significantly enriched GO terms. Statistical

significance in the enrichment of biological process GO terms (=102) was numerically assessed by comparing with 10 000 equally sized random samples of genes for up-regulated differentially expressed genes in the liver. The most significantly enriched categories retrieved are shown in order of importance from top to bottom. Significance threshold (Adjusted P-value < 0.05) was adjusted for multiple testing by Benjamini-Hochberg correction.

Table 6.4. Shared GO terms between the liver and each of the subcortical areas and the cortex of all the developmental stages. * has one same identical GO term.

Differentially expressed genes	Tissues	Number of shared GO term
Down-regulated	Subcortical areas and liver	14*
Up-regulated	Subcortical areas and liver	1
Up-regulated	Cortex and liver	3*
Down-regulated	Cortex and liver	0

Table 6.5. GOIDs and GO term. Shared GO term and GOIDs between the liver and the subcortical areas in significantly differentially expressed down- and up-regulated genes and between the liver and the cortex in significantly differentially expressed up-regulated genes. * GO:0008285 is identical GO term.

GOID	GO term	Tissue	Differentially expressed
GO:0006810	transport	subcortical areas&liver	down-regulated genes
GO:0006811	ion transport	subcortical areas&liver	down-regulated genes
GO:0006814	sodium ion transport	subcortical areas&liver	down-regulated genes
*GO:0008285	negative regulation of cell proliferation	subcortical areas&liver	down-regulated genes
GO:0009612	response to mechanical stimulus	subcortical areas&liver	down-regulated genes
GO:0009725	response to hormone	subcortical areas&liver	down-regulated genes
GO:0009749	response to glucose	subcortical areas&liver	down-regulated genes
GO:0030154	cell differentiation	subcortical areas&liver	down-regulated genes
GO:0032355	response to estradiol	subcortical areas&liver	down-regulated genes
GO:0035264	multicellular organism growth	subcortical areas&liver	down-regulated genes
GO:0035725	sodium ion transmembrane transport	subcortical areas&liver	down-regulated genes
GO:0042493	response to drug	subcortical areas&liver	down-regulated genes
GO:0043010	camera-type eye development	subcortical areas&liver	down-regulated genes
GO:0055085	transmembrane transport	subcortical areas&liver	down-regulated genes
GO:0001666	response to hypoxia	subcortical areas&liver	up-regulated genes
*GO:0008285	negative regulation of cell proliferation	cortex&liver	up-regulated genes
GO:0010628	positive regulation of gene expression	cortex&liver	up-regulated genes
GO:0033138	positive regulation of peptidyl-serine phosphorylation	cortex&liver	up-regulated genes

Discussion

Grb10 is an imprinted gene in human and mice with a tissue-specific manner expression (Arnaud et al., 2003; Hikichi et al., 2003; Monk et al., 2009). In the mouse, maternal allele expressed in most of the tissue while paternal allele expressed in certain parts of the brain (Garfield et al., 2011). Mutation of Grb10 from maternal side leads to overgrowth of the embryo including disproportion at the level of tissues and organs (Smith et al., 2007) while from paternal side leads to dominant behaviour in adult mice (Garfield., 2011).

In this chapter, I compared differentially expressed gene sets of maternal mutant Grb10KO$^{m/+}$ mice and paternal mutant Grb10KO$^{+/p}$ when compared to corresponding wild-type in order to assess whether changes in transcriptional profiles in subcortical areas are specific to the brain or instead part of a systematic effect of the depletion of Grb10.

FastQC analysis showed good quality of the reads which facilitate the annotation and quantification of all genes in the samples. The overall observation on the amount of differentially expressed genes of both up- and down-regulated in the liver of E18.5 has been detected 279 which are higher than what have been detected in the subcortical areas and the cortex in all the developmental stages (139 genes, 26 genes, respectively). This is may be due to molecular changes in brain in these mutant mice might be more subtle than those seen in tissues with maternal allele expression. However, it may also be the case that more generally, transcriptional shifts might be generally more extreme in the liver compared to the brain. For example, Masuo et al., 2009 found that in adult alcoholic rat the differential gene expression in liver was higher than the brain. Similar result was observed in hypoxia-exposed medaka (Ju et al., 2007).

As in the case of the brain, analyses of the Grb10KO liver samples showed residual presence of Grb10 transcripts, although with a significant reduction in expression compared to the wild-type. This suggests that β-geo insertion does not completely prevent Grb10 expression. Whether functional proteins are translated from these transcripts is unknown. Assessment of the overlap in the sets of differentially expressed genes when comparing liver and the subcortical brain tissue, as well as the neocortex (Figure 6.3 and Figure 6.4), suggests that most of the differentially

expressed genes are specific for single tissue and only few genes are downregulated in both the maternal and the paternal KO mice.

Gene ontology enrichment analysis conducted to gain further knowledge into the functional coherence of the differentially expressed genes by recognizing specific targets of biological processes. Identifying few shared GO terms among the liver and the brain tissue (Table 6.4) suggest that most of the differentially expressed genes are regulated in a spatially specific manner, while few might share the same function.

I further examined alternative splicing variants in the liver samples (Table 6.2). I was able to detect these splice variants of Grb10 transcript in which one of them mGrb10-206 has been specifically transcript in the liver of E18.5 mice as the other 2 transcripts (mGrb10-2034 and mGrb10-201) have been detected in the brain tissue at certain developmental stages other than E18.5 which confirm a tissue-specific manner transcription of Grb10 (Charalambous et al., 2003; Garfield et al., 2011). It also suggests that certain splice variant of Grb10 might have parental-specific manner transcription. tissue-specific expression of the transcript splice variants. Furthermore, it confirmed that $Grb10^{m/+}$ had no effect on Grb10 expression in the brain (Wang et al., 2007).

Chapter VII Final Discussion

Introduction

Several studies of the human brain transcriptome have shed light on how the different brain structures, ages, gender and cell types differ in their specific gene expression patterns (Oldham et al., 2008; Weickert et al., 2009; Kang et al., 2011; Darmanis et al., 2015). The brain structure networks of vertebrate species show a high level of similarity in regulating the social behaviour (Newman, 1999; Goodson, 2005) that shows a common conserved pathway (O'Connell & Hofmann, 2011) and the variation of gene expression in the brain may play a critical role in determining behavioural phenotypes.

Grb10 has to date a somewhat unique imprinted expression pattern with a tissue-specific manner expression (Arnaud et al., 2003; Hikichi et al., 2003; Monk et al., 2009) it has been detected in almost every tissue in the mouse embryo including the brain, as the maternal allele expressed in most peripheral tissues while paternal allele expressed in certain parts of the brain (Charalambous et al., 2003; Hikichi et al., 2003; Garfield et al., 2011) which contribute to the behaviour (social dominant in mice) (Garfield et al., 2011) and brain growth (Garfield, 2007).

To study the molecular basis of this social behaviour, I started by examining existing variations of Grb10 expression profiles across different developmental stages (E18.5, 1W, 1M, 3M, 6M) in the subcortical areas and the cortices. The knockout line used for these samples is that reported by Garfield et al. (2011).

Grb10 was detected in all those stages, however, knocking line for these mice was not successfully established it didn't cease or suppress Grb10 expression due to presence of Grb10 in the subcortical areas of mutant mice with lower level expression compared with WT (Table 3.1 and Table 3.3). Besides, this gene also detected in the cortex of mutant mice with lower level compared to the subcortical areas (Table 3.2 and Table 3.4) which confirm that knocking line for these mice didn't cease Grb10 expression. In addition, it shows that Grb10 also expressed in the cortex but in much more lower levels.

Another possibility of detecting Grb10 is that those were maternal Grb10 as in late embryonic development limited expression of maternal allele was found in ventricular ependymal layers, epithelium of choroid plexus and meninges of the brain (Garfield

et al., 2011), it might also have expressed in other developmental stages. Although the expression of Grb10 in the cortex is undetectable or expressed weakly (Cowley, 2009).

Cowley et al. (2014) also described two mouse models of Grb10 ablation, which are generated by the integration of a *LacZ* reporter gene-trap cassette in different loci of Grb10KO$^{+/p}$ (Garfield et al., 2011) and Grb10Δ2-4$^{+/p}$ (Charalambous et al., 2003). That shows identical phenotypic consequences but differ in expression site of the brain. Further investigations must be performed to elucidate the type of the Grb10 allele has been detected as the paternal allele arises from a different promoter region than the maternal (Arnaud et al., 2003; Plasschaert & Bartolomei, 2015).

In order to comprehensively interpret and shed light on the potential molecular interactors of the paternal Grb10 allele in the brain during development and adult stages the transcriptome profile analyses of the Grb10 paternal KO$^{+/p}$ using raw RNA-Seq datasets was made.

Analysis of subcortical areas and cortex through a series of developmental stages showed transcriptional alterations pattern differs in each developmental stage as the differentially expressed genes of both up- and down-regulated can be high in one developmental stage and less in other. Most transcriptional alteration n =256 (Table 4.3), highest number of differentially expressed genes =139 (Figure 3.7 and Figure 3.8) and GO enrichment targeted the biological process associated with the social dominant behaviour in mutant Grb10KO$^{+/p}$ mice (GO:0007625) (Figure 3.11A and supplementary Figure S3.1) appeared in module of down-regulated differentially expressed genes in the subcortical areas confirmed that mutation of Grb10 which leads to social dominant can play a role in those alterations.

Most of the cellular, physiological and developmental functions are the result of gene groups interacting and cooperating instead of individual genes acting in isolation (Hartwell et al., 1999). Clusters of genes involved in the same pathway, biological process tend to display correlated expression patterns reflecting their functional associations (Eisen et al., 1998; Homouz & Kudlicki, 2013; Marco et al., 2009).

For this study, common alternative splicing transcripts across 2 different tissues of the brain (subcortical areas and cortex) at different stages were analysed. My results by

using SplAdder shows difficulty in distinguishing Grb10 alternative splicing events from RNA-Seq data may be because of its algorithm or due to that some alternatively spliced mRNA forms are miscategorised as genomic DNA, causing to be excluded by the procedure for analysing its splicing. Cross-hybridization in which RNA hybrids by complementary base pairing between two molecules that are not identical in sequence could also give false result when sequence similar genes have strong tissue-specific regulation (Relógio et al., 2005). However, using the edgeR software with Kallisto output showed that highest number of significantly differentially expressed (FDR < 0.05) splice variant transcripts appeared in the subcortical areas at age of 3 months old mice (Table 4.3) which may be due to highest number that appeared at this stage (Figure 3.7). Different transcript splice variant at different stages may be due to regulatory reorganization model where the developmental gene expression programme undergoes an overall reassembly of gene regulatory interactions (Monzón-Sandoval et al., 2016).

The development of the nervous system is an extremely dynamic, elaborated and complex process and coordination of the gene expression is crucial. In this study, I asked whether the architecture of the underlying regulatory network in Grb10 mutant mice in the subcortical areas and the cortex at different stages is a constant or variable feature of the developmental programme. Changes in gene co-expression effectively reveal events of deregulation and dysfunction affecting specific certain biological processes (Choi et al., 2005) and changes in co-expression have been described in the context of age-related changes (Southworth et al., 2009).

Gene network approaches have been used to study a variety of biological system by exploring the observed relationships between gene products. Furthermore, gene co-expression analysis can be used to identify genes with similar expression patterns. It can establish a wide range of biological information such as biological pathways, cell type expression and shared gene functions (Kang et al., 2011; Hawrylycz et al., 2012; Grange et al., 2014).

Differentially co-expressed genes network analysis was carried out to detect any relation between sets of genes and gene regulatory networks in the subcortical areas and the cortex triggered by a mutation of Grb10KO$^{+/p}$. The gene network revealed that the transcriptome in the subcortical areas and the cortex are organized into discrete clusters each involved in the regulatory defined sets of biological functions.

The results revealed differentially co-expressed gene clusters integrated by genes preferentially associated to particular biological processes in which each of the clusters consists of genes involved in specific with rare or very few functional overlap between modules sets of functions. However, the cluster module where the Grb10 is presented has been shown in the subcortical areas (dark red module) and the cortex module (yellow module) but Bonferroni adjusted differential co-expression p-value < 0.001 has been observed only in dark red cluster module (Table 5.1) this can suggest that Grb10 is predominant in the subcortical areas where it interact with other molecular regulators or substances to govern the gene expression levels of mRNA mediating the phenotype such as the social dominance.

Maternal mutant Grb10KO$^{m/+}$ mice have been used to compare their results to the paternal mutant Grb10KO$^{+/p}$ mice. The most interesting finding was that the transcriptome level appears to have different regional distributions in the brain tissue and at different developmental stages (Table 4.3) and the liver of E18.5 (Table 6.1), these differences reflect developmental and functional differences between in brain regions (paternal allele) and other tissue organs as liver (maternal allele). Furthermore, shared GO term was highest (=14) in the subcortical areas with differentially expressed down-regulated genes (Table 6.4) although this could be because the highest number of genes was detected at 3 months old mice.

In additional of that at E18.5 mice specifically different Grb10 transcript splice variants were detected in the liver and the subcortical areas while no Grb10 transcript splice variant was detected in the neocortex. which is obvious because of the impact of a maternal and paternal allele on specific tissue. Also, since Grb10 gene is expressed with different splice variants, it is possible that depending on the tissue-specific abundance of the different splice variants, Grb10 regulates receptor downstream signalling differentially. Moreover, differential gene expression in the module of the biological process showed several GO terms associated with behaviour including grooming behaviour (GO:0007625) which distinctly related to the Grb10KO brain region in the subcortical areas rather than in the cortex or the liver.

Future work:

It will be desirable to carry out the tube test as a behavioural test of dominance.
In this test as described by Garfield et al. (2011) is carried out between socially isolated unfamiliar adult knock-out mice with WT mice both matched for body size. The tube test was first described in 1961 as a measure of social dominance (Lindzey et al., 1961) and consists of placing mice at either end of a clear tube and then observed which mice backs off from the tube during these forced encounters. The paternal mice Grb10KO$^{+/p}$ will significantly less likely to back down than WT.

The sample size is an important element in the design of research projects, it is not clear an exact number of biological replicates are needed to assure valid biological interpretation of the results. The larger sample size is more representative of the population and provides enough statistical power for meaningful analysis that can produce artefactual data. It limits the influence of outliers or extreme observations and produces results among variables that are significantly different, although the economics may play roles in it.

The common misconception about sampling in qualitative research is that numbers are unimportant in ensuring the adequacy of a sampling strategy (Sandelowski, 1995).
In any experiment requiring human or animal, the sample size is the main issue for ethical reasons, in addition, the sample size is not always the main issue; it is only one aspect of the quality of a study design (Lenth, 2001). Bioinformatic tools also important for judging the sample size. For instance, Schurch et al. (2016) found that due to failure of some of bioinformatic tools using RNA-Seq data to control their FDR adequately for detecting medium to high fold change of differential expressed genes he suggested at least six biological replicates instead of three which is a minimum number used to have a standard deviation.

In my opinion, the RNA-Seq data of replicates (=3) which had been used although showed a significant number of differentially expressed genes that expressed widely in tissues when, in reality, their profile may be much more limited. Having additional replicates might make it clearer and purposefully. On top of that in the present study, I used 5 developmental stages and the last stage was 6 months old mice having elder stages than that for instance 12 months old mice will give more

noticeable information.

Initially, it will be important to ascertain if the transcriptional terminator is confirmed that this mouse is truly knocked out that's why it will be desirable to generate further Grb10 knock mice targeted ES cell line to ascertain that Grb10 will not be detected in RNA-Seq data. For example, the researchers cannot assure the exact moment when the microorganism (e.g. virus) develops; they can just determine is that the virus developed before or after the test.

To assure that Grb10 is completely knock out at all the developmental stages not only during the dominant phenomena has been observed at a certain stage, a new technique called Conditional gene knockout can be used. This technique allows much more sophisticated experiments than traditional knockout, it used to eliminate a specific gene in a certain tissue at specific times rather than being deleted from the beginning of life. Researchers use certain enzymes at site-specific recombinase systems (the enzymes catalysing site-specific recombination) such as Cre-loxP (Hoess et al., 1982; Orban et al., 1992), Flp-FRT (Sternberg et al., 1981), and ΦC31 (Thyagarajan et al., 2001) within particular nucleotide sequences that specifically cleaves DNA at consensus sequences to engineer tissue-specific knockout mice. These conditional knockout mice enable researchers to analyse the effect of the gene's absence in a specific lineage of cells (Clarke, 2000; Guan et al., 2010; Friedel et al., 2011; Zhang et al., 2012B). Moreover, generating samples from different stages of maternal knock-out mice $Grb10KO^{m/+}$ to understand the regulation mechanism of transcripts alternative splicing with other genomic analysis at each stage and compare them with paternal knock-out mice $Grb10KO^{+/p}$.

Obviously, the bioinformatic method is a powerful tool, but it does have its limitations. These limitations are based on the fact that a result must be testable and falsifiable, and that experiments and observations be repeatable.

Understanding the limitations of RNA-Seq is critical for accurate biological interpretations, as one important limitation of the widely used RNA-Seq approaches for studying RNAs particularly small RNAs is their inability to provide an absolutely quantitative view of these transcripts as the longer transcripts are broken into more fragments than are shorter ones resulting in a biased estimation of expression levels (Ozsolak & Milos, 2011; Tandonnet & Torres, 2017). According to Oshlack & Wakefield (2009) to minimize this bias, the levels of expression can be corrected

by the size of the transcript or by using the contig size from the de novo reconstruction of the transcript. Nevertheless, this correction does not entirely solve the problem owing to the transcript size, as the sampling is higher for longer transcripts. Other limitation of using one set RNA-Seq can be Transcript quantitation can be affected by biases introduced during cDNA library construction and sequence alignment in addition to lack of standardization between sequencing platforms and read depth, equivalent to the percentage of total transcripts sequenced, can compromise reproducibility (Whitley et al., 2016). In addition to biocomputational parametric programs and algorithm limitations.
In spite of these limitation, the current applications and information provided by RNA-Seq are enough to make it valuable those are determination of transcriptome profile and gene expression levels.

To further verify the RNA-Seq results of assessing Grb10 expression in different tissue and developmental stages different methods can be conducted such as a quantitative polymerase chain reaction (qPCR) which is based on the analysis of RNA to quantify gene expression and identify biodegradation activity. Also, in situ hybridization can be conducted to check the presence of primers specific to Grb10 and the inserted β-geo gene-trap cassette at different developmental stages that gain information of expression pattern of Grb10 by visualizing the target-specific probe within the brain tissue samples. In addition, paternal allele arises from a different promoter region than the maternal (Arnaud et al., 2003; Plasschaert & Bartolomei, 2015), therefore, performing in situ hybridization can detect the presence of primers specific to maternal Grb10 which can give us an idea if the maternal allele is expressed in those brain regions.

In conclusion:

Grb10 is a unique imprinted gene in human and mice with a tissue-specific manner expression. It plays role in growth and dominant behaviour in adult mice.

In this project, I aimed to uncover the potential molecular interactors related to the Grb10KO$^{+/p}$ compared to WT in the subcortical areas and the cortex of the mice at five developmental stages (E18.5, 1 week old, 1 month old, 3 months old and 6 months old mice) and compare the patterns of differential gene expression with the

changes observed associated with the maternal Grb10KO$^{m/+}$ transcriptome profiles in the liver.

For this, I obtained whole genome transcriptional profiles using Illumina RNA-Seq technology on these samples, using bioinformatic tools to annotate the genes and their alternative splicing events and analyses of differential gene expression. I identified a set of genes altered in the paternal mutant Grb10KO$^{+/p}$ and how these sets of genes change over time. I also, carry out differentially co-expressed genes network analysis to detect any relation between sets of genes and gene regulatory networks in the subcortical areas and the cortex triggered by a mutation of Grb10KO$^{+/p}$.

I found that the highest number of differentially expressed genes appeared at adult stage of 3 months old mice where Grb10 is down-regulated in the subcortical areas. The GO categories confirmed the dominant behaviour in biological process for down-regulated differentially expressed genes in the subcortical areas. However, GO enrichment for the overlapped genes in the subcortical areas and the cortex showed no adjusted significant values. Moreover, no considerable amount of differentially expressed genes shared across the developmental groups in both brain tissue except 1-3 genes. In addition, no big differences found at a total number of alternative splicing events, and the number of alternative splicing events within a gene between the subcortical areas and the cortex.

Co-expression network analyses revealed significant shifts in gene to gene relationships in the KO context.

Finally, different parental alleles are associated with different sets of genes in different tissues which is confirmed by distinct gene expression profile changes in the Grb10KO in the brain from gene expression in the liver.

Although the significance of this study is clear, more research is required as discussed earlier in the future work.

References

Amar, D., Safer, H., Shamir, R., 2013. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput Biol.* 9 (3), pp.e1002955.

Anders, S., and Huber, W., 2010. DESeq: Differential expression analysis for sequence count data. *Genome Biology*. 11 (10), pp.R106.

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq - a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 31, pp.166-9.

Anders, S., Reyes, A., Huber, W., 2012. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 22 (10), pp.2008-17.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. Available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed date 8th March 2017].

Arnaud, P., Monk, D., Hitchins, M., Gordon, E., Dean, W., Beechey, C.V., Peters, J., Craigen, W., Preece, M., Stanier, P., Moore, G.E., Kelsey, G., 2003. Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. *Human Molecular Genetics*. 12 (9), pp.1005-19.

Arrant, A.E., Filiano, A.J., Warmus, B.A., Hall, A.M., Roberson, E.D., 2016. Progranulin haploin sufficiency causes biphasic social dominance abnormalities in the tube test. *Genes Brain Behav*. 15 (6), pp.588-603.

Bai, R.Y., Jahn, T., Schrem, S., Munzert, G., Weidner, K.M., Wang, J.Y. and Duyster, J., 1998. The SH2-containing adapter protein GRB10 interacts with BCR-ABL. *Oncogene*. 17 (8), pp.941-8.

Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., Mardis, E.R., Sadar, M.D., Siddiqu,

A.S., Marra, M.A., Jones, S.J., 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics*. 7, pp.1-11.

Barbaux, S., Gascoin-Lachambre, G., Buffat, C., Monnier, P., Mondon, F., Tonanny, M.B., Pinard, A., Auer, J., Bessières, B., Barlier, A., Jacques, S., Simeoni, U., Dandolo, L., Letourneur, F., Jammes, H. and Vaiman, D., 2012. A genome-wide approach reveals novel imprinted genes expressed in the human placenta. *Epigenetics*. 7 (9), pp.1079-90.

Barbosa-Morais, N.L., Irimia, M., Pan, Q., et al. (17 co-authors)., 2012. The evolutionary landscape of alternative splicing in vertebrate species. *Science.* 338 (6114), pp.1587-93.

Beck, S., Penque, D., Garcia, S., Gomes, A., Farinha, C., Mata, L., Gulbenkian, S., Gil-Ferreira, K., Duarte, A., Pacheco, P., Barreto, C., Lopes, B., Cavaco, J., Lavinha, J., Amaral, M.D., 1999. Cystic fibrosis patients with the 3272-26A-->G mutation have mild disease, leaky alternative mRNA splicing, and CFTR protein at the cell membrane. *Hum Mutat*. 14 (2), pp.133-44.

Beffert, U., Weeber, E.J., Durudas, A., Qiu, S., Masiulis, I., Sweatt, J.D., Li, W.P., Adelmann, G., Frotscher, M., Hammer, R.E., Herz, J., 2005. Modulation of synaptic plasticity and memory by Reelin involves differential splicing of the lipoprotein receptor Apoer2. *Neuron*. 47 (4), pp.567-79.

Belgard, T.G., Marques, A.C., Oliver, P.L., Abaan, H.O., Sirey, T.M., Hoerder-Suabedissen, A., García-Moreno, F., Molnár, Z., Margulies, E.H., Ponting, C.P., 2011. A transcriptomic atlas of mouse neocortical layers. *Neuron*. 71 (4), pp.605-16.

Ben-Dov, C., Hartmann, B., Lundgren, J. & Valcarcel, J., 2008. Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem.* 283 (3), pp.1229-33.

Bhora, F.Y., Dunkin, B.J., Batzri, S., Aly, H.M., Bass, B.L., Sidawy, A.N., Harmon, J.W., 1995. Effect of Growth Factors on Cell Proliferation and Epithelialization in Human Skin. *J Surg Res*. 59 (2), pp.236-44.

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J. and Nekrutenko, A., Galaxy Team., 2010. Manipulation of FASTQ data with Galaxy. *Bioinformatics*. 26 (14), pp.1783-5.

Blencowe, B.J., 2006. Alternative splicing: new insights from global analyses. *Cell.* 126 (1), pp.37-47.

Bray, N., Pimentel, H., Melsted, P. and Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*. 34 (5), pp.525-7.

Bullard, J.H., Purdom, E., Hansen, K.D., Dudoit, S., 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*. 11, pp.94.

Cao, X.R., Lill, N.L., Boase, N., et al. (11 co-authors)., 2008. *Nedd4* controls animal growth by regulating IGF-1 signaling. *Sci Signal*. 1 (38), pp.ra5.

Carpenter, A.E. and Sabatini, D.M., 2004. Systematic genome-wide screens of gene function. *Nat Rev Genet.* 5 (1), pp.11-22.

Charalambous, M., Cowley, M., Geoghegan, F., Smith, F.M., Radford, E., Hurst, L.D.H. and Ward, A., 2010. Maternally-inherited Grb10 reduces placental size and efficiency. *Dev Biol*. 337, pp.1-8.

Charalambous, M., Smith, F.M., Bennett, W.R., Crew, T.E., Mackenzie, F. and Ward, A., 2003. Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. *Proc Natl Acad Sci.* 100 (14), pp.8292-7.

Chen, C., Cheng, L., Grennan, K., Pibiri, F., Zhang, C., Badner, J.A., Gershon, E.S., Liu, C., 2013. Two gene co-expression modules differentiate psychotics and controls. *Mol Psychiatry*. 18 (12), pp.1308-14.

Chen, M. and Manley, J.L., 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.* 10 (11), pp.741-54.

Ching, T., Huang, S., and Garmire, L.X., 2014. Power analysis and sample size estimation for RNA-Seq differential expression. *RNA*. 20 (11), pp.1684-96.

Choi, J.K., Yu, U., Yoo, O.J., Kim, S., 2005. Differential coexpression analysis using microarray data and its application to human cancer. *Bioinformatics.* 21 (24), pp.4348-55.

Clark, F. and Thanaraj, T.A., 2002. Categorization and characterization of transcript-confirmed constitutively and alternatively spliced introns and exons from human. *Hum Mol Genet*. 11 (4), pp.451-64.

Clark, T.A., Sugnet, C.W., Ares, M. Jr., 2002. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science.* 296 (5569), pp.907-10.

Clarke, A.R., 2000. Manipulating the germline: its impact on the study of carcinogenesis. *Carcinogenesis*. 21 (3), 435-41.

Cloonan, N., Forrest, A.R., Kolle, G., et al. (19 co-authors)., 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods.* 5 (7), pp.613-9.

Colley, B.S., Cavallin, M.A., Biju, K., Marks, D.R., Fadool, D.A., 2009. Brain-derived neurotrophic factor modulation of Kv1.3 channel is disregulated by adaptor proteins Grb10 and nShc. *BMC Neurosci*. 10, pp.8.

Cook, K.K. and Fadool, D.A., 2002. Two adaptor proteins differentially modulate the phosphorylation and biophysics of Kv1.3 ion channel by SRC kinase. *J Biol Chem*. 277 (15), pp.13268-80.

Cooper, T.A. and Mattox, W., 1997. The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet*. 61 (2), pp.259-66.

Cowley, M.A., 2009. *Investigating the role of the imprinted Grb10 gene in the regulation of maternal nutrient transfer.* Thesis (PhD.). University of Bath.

Cowley, M., Garfield, A.S., Madon-Simon, M., Charalambous, M., Clarkson, R.W., Smalley, M.J., Kendrick, H., Isles, A.R., Parry, A.J., Carney, S., Oakey, R.J., Heisler, L.K., Moorwood, K., Wolf, J.B. and Ward, A., 2014. Developmental Programming Mediated by Complementary Roles of Imprinted Grb10 in Mother and Pup. *PLoS Biol.* 12 (2), pp.e1001799.

Curley, J.P., 2011. Is there a genomically imprinted social brain?. *Bioessays*. 33 (9), pp.662-8.

Daly, J.R., 1998. The Grb7 family of signaling proteins. *Cell Signal.* 10 (9), pp.613-8.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., et al., 2015. A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Science of the United States of America*. 112 (23), pp.7285-90.

de Esch, C.E., van den Berg, W.E., Buijsen, R.A., Jaafar, I.A., Nieuwenhuizen-Bakker, I.M., Gasparini, F., Kushner, S.A., Willemsen, R., 2015. Fragile X mice have robust mGluR5-dependent alterations of social behaviour in the Automated Tube Test. *Neurobiol Dis.* 75, pp.31-9.

de la Fuente, A., 2010. From 'differential expression' to 'differential networking' – identification of dysfunctional regulatory networks in diseases. *Trends Genet.* 26 (7), pp.326-33.

de Jong, S., Boks, M.P., Fuller, T.F., Strengman, E., Janson, E., de Kovel, C.G., Ori, A.P., Vi, N., Mulder, F., Blom, J.D., 2012. A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. *PLoS One*. 7 (6), pp.e39498.

Deng, Y., Zhang, M. and Riedel, H., 2008. Mitogenic roles of Gab1 and Grb10 as direct cellular partners in the regulation of MAP kinase signalling. *J Cell Biochem.* 105 (5), pp.1172-82.

Depetris, R.S., Hu, J., Gimpelevich, I., Holt, L.J., Daly, R.J., Hubbard, S.R., 2005. Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. *Mol Cell*. 20 (2), pp.325-33.

Depetris, R.S., Wu, J. and Hubbard, S.R., 2009. Structural and functional studies of the Ras-associating and pleckstrin-homology domains of Grb10 and Grb14. *Nat Struct Mol Biol*. 16 (8), pp.833-9.

Desbuquois, B., Carré, N., Burnol, A.F., 2013. Regulation of insulin and type 1 insulin-like growth factor signaling and action by the Grb10/14 and SH2B1/B2 adaptor proteins. *FEBS J.* 280 (3), pp.794-816.

Dey, B.R., Frick, K., Lopaczynski, W., Nissley, S.P., Furlanetto, R.W., 1996. Evidence for the direct interaction of the insulin-like growth factor I receptor with IRS-1, Shc, and Grb10. *Mol Endocrinol.* 10 (6), pp.631-41.

Dong, L.Q., Du, H., Porter, S.G., Kolakowski, L.F. Jr., Lee, A.V., Mandarino, L.J., Fan, J., Yee, D., Liu, F., 1997. Cloning, chromosome localization, expression, and characterization of an Src homology 2 and pleckstrin homology domain-containing insulin receptor binding protein hGrb10gamma. *J Biol Chem.* 272 (46), pp.29104-12.

Drewe, P., Stegle, O., Hartmann, L., Kahles, A., Bohnert, R., Wachter, A., Borgwardt, K., Rätsch, G., 2013. Accurate detection of differential RNA processing. *Nucleic Acids Res.* 41 (10), pp.5189-98.

Dufresne, A.M. and Smith, R.J., 2005. The adapter protein GRB10 is an endogenous negative regulator of insulin-like growth factor signaling. *Endocrinology*. 146 (10), pp.4399-409.

Dunbar, R.I., 2010. The social role of touch in humans and primates: behavioural function and neurobiological mechanisms. *Neurosci Biobehav Rev*. 34 (2), pp.260-8.

ENSEMBL, 2017. [Online]. Available from http://www.ensembl.org [Accessed date 21th January 2017].

Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A.* 95 (25), pp.14863-8.

Feil, R. and Berger, F., 2007. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet*. 23 (4), pp.192-9.

Filiano, A.J., Martens, L.H., Young, A.H., Warmus, B.A., Zhou, P., Diaz-Ramirez, G., Jiao, J., Zhang, Z., Huang, E.J., Gao, F.B., Farese, R.V. Jr. & Roberson, E.D., 2013. Dissociation of frontotemporal dementia–related deficits and neuroinflammation in progranulin haploinsufficient mice. *J Neurosci.* 33 (12), pp.5352-61.

Frantz, J.D., Giorgetti-Peraldi, S., Ottinger, E.A. and Shoelson, S.E., 1997. Human GRB-IRbeta/GRB10. Splice variants of an insulin and growth factor receptor-binding protein with PH and SH2 domains. *J Biol Chem* 272 (5), pp.2659-67.

Friedel, R.H., Wurst, W., Wefers, B., Kühn, R., 2011. Generating Conditional Knockout Mice. *Methods in Molecular Biology.* 693, pp.205-31.

Gaiteri, C., Ding, Y., French, B., Tseng, G.C., Sibille, E., 2014. Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders. *Genes Brain Behav*. 13 (1), pp.13-24.

Galaxy (version 0.67) [computer program]. Available from: http://usegalaxy.org [Accessed 16 January 2016].

Gao, Q., Ho, C., Jia, Y., Jia, Y., Li, J.J., Huang, H., 2012. Biclustering of linear patterns in gene expression data. *J Comput Biol.* 19 (6), pp.619-31.

Garber, M., Grabherr, M.G., Guttman, M., Trapnell, C., 2011. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat Methods*. 8 (6), pp.469-77.

Garcia-Blanco, M.A., Baraniak, A.P., Lasda, E.L., 2004. Alternative splicing in disease and therapy. *Nat Biotechnol.* 22 (5), pp.535-46.

Garfield, A.S., 2007. *Investigating the roles of mouse Grb10 in the regulation of growth and behaviour*. Thesis (PhD.). University of Bath.

Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D., Wilkinson, L.S., Isles, A.R. and Ward, A., 2011. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature.* 469 (7331), pp.534-8.

Gierliński, M., Cole, C., Schofield, P., Schurch, N.J., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G., OwenHughes, T., Blaxter, M., and Barton, G.J., 2015. Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment. *Bioinformatics*. 31 (22), pp.3625-3630.

Giorgetti-Peraldi, S., Murdaca, J., Mas, J. C. and Van Obberghen, E., 2001. The adapter protein, Grb10, is a positive regulator of vascular endothelial growth factor signalling. *Oncogene*. 20 (30), pp.3959-68.

Goecks, J., Nekrutenko, A., Taylor, J., Galaxy Team., 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the lifesciences. *Genome Biol*. 11 (8), pp.R86.

Goodson, J.L., 2005. The vertebrate social behaviour network: evolutionary themes and variations. *Horm Behav.* 48 (1), pp.11-22.

Grange, P., Bohland, J.W., Okaty, B.W., Sugino, K., Bokil, H., Nelson, S.B., Ng, L., Hawrylycz, M., Mitra, P.P., 2014. Cell-type-based model explaining coexpression patterns of genes in the brain. *Proc Natl Acad Sci USA*. 111 (14), pp.5397-402.

Grant, G.R., Liu, J., Stoeckert, C.J. Jr., 2005. A practical false discovery rate approach to identifying patterns of differential expression in microarray data. *Bioinformatics*. 21 (11), pp.2684-90.

Graveley, B.R., 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet*. 17 (2), pp.100-7.

Gruppuso, P.A., Boylan, J.M., Vaslet, C.A., 2000. Identification of candidate growth-regulating genes that are overexpressed in late gestation fetal liver in the rat. *Biochim Biophys Acta.* 1494 (3), pp.242-7.

Guan, C., Ye, C., Yang, X., and Gao, J., 2010. A review of current large-scale mouse knockout efforts. *Genesis*. 48 (2), pp.73-85.

Harris, M.A., Clark, J., Ireland, A., et al. (59 co-authors)., 2004. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*. 32 (Database issue), pp.D258-61.

Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W., 1999. From molecular to modular cell biology. *Nature*. 402 (6761 Suppl), pp.C47-52.

Hatayama, M., Ishiguro, A., Iwayama, Y., et al. (11 co-authors)., 2011. Zic2 hypomorphic mutant mice as a schizophrenia model and ZIC2 mutations identified in schizophrenia patients. *Sci Rep*. 1, pp.16.

Han, D.C., Shen, T.L., Guan, J.L., 2001. The Grb7 family proteins: structure, interactions with other signalling molecules and potential cellular functions. *Oncogene.* 20 (44), pp.6315-21.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P., Vidal, M., 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*. 430 (6995), pp.88-93.

Han, X., Wu, X., Chung, W.Y., Li, T., Nekrutenko, A., Altman, N.S., Chen, G., Ma, H., 2009. Transcriptome of embryonic and neonatal mouse cortex by high-throughput RNA sequencing. *Proc Natl Acad Sci U S A*. 106 (31), pp.12741-6.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., et al., 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 489 (7416), pp.391-9.

He, W., Rose, D.W., Olefsky, J.M., Gustafson, T.A., 1998. Grb10 Interacts Differentially with the Insulin Receptor, Insulin-like Growth Factor I Receptor, and Epidermal Growth Factor Receptor via the Grb10 Src Homology 2 (SH2) Domain and a Second Novel Domain Located between the Pleckstrin Homology and SH2 Domains. *J Biol Chem*. 273 (12), pp.6860-7.

Hikichi, T., Kohda, T., Kaneko-Ishino, T. and Ishino, F., 2003. Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. *Nucleic Acids Res.* 31 (5), pp.1398-1406.

Hitchins, M.P., Monk, D., Bell, G.M., Ali, Z., Preece, M.A., Stanier, P. and Moore, G.E., 2001. Maternal repression of the human GRB10 gene in the developing central nervous system; evaluation of the role for GRB10 in Silver-Russell syndrome. *Europ. J. Hum. Genet.* 9 (2), pp.82-90.

Hiller, D., Wong, W.H., 2013. Simultaneous isoform discovery and quantification from RNA-Seq. *Stat Biosci*. 5 (1), pp.100-18.

Holt, L.J. and Siddle, K., 2005. Grb10 and Grb14: enigmatic regulators of insulin action—and more?. *Biochem. J.* 388 (Pt 2), pp.393-406.

Homouz, D., Kudlicki, A.S., 2013. The 3D organization of the yeast genome correlates with co-expression and reflects functional relations between genes. *PLoS One*. 8 (1), pp.e54699.

Hoess, R.H., Ziese, M., Sternberg, N., 1982. P1 site-specific recombination: nucleotide sequence of the recombining sites. *PNAS.* 79 (11), pp.3398-402.

Hsu, P.P., Kang, S.A., Rameseder, J., Zhang, Y., Ottina, K.A., Lim, D., Peterson, T.R., Choi, Y., Gray, N.S., Yaffe, M.B., Marto, J.A. and Sabatini, D.M., 2011. The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signalling. *Science*. 332 (6035), pp.1317-22.

Hu, R., Qiu, X., Glazko, G., Klebanov, L., Yakovlev, A., 2009. Detecting intergene correlation changes in microarray analysis: a new approach to gene selection. *BMC Bioinformatics*. 10, pp.20.

Hu, Y., Huang, Y., Du, Y., Orellana, C.F., Singh, D., Johnson, A.R., et al. (18 co-authors)., 2013. DissSplice: the genome-wide detection of differential splicing events with RNA-seq. *Nucleic Acids Res*. 41 (2), pp.e39.

Hu, Z.Q., Zhang, J.Y., Ji, C.N., Xie, Y., Chen, J.Z., Mao, Y.M., 2010. Grb10 interacts with Bim L and inhibits apoptosis. *Mol Biol Rep*. 37 (7), pp.3547-52.

Huang, Q. and Szebenyi, D.M., 2010. Structural basis for the interaction between the growth factor-binding protein GRB10 and the E3 ubiquitin ligase *NEDD4*. *J Biol Chem.* 285 (53), pp.42130-9.

Hudson, N.J., Reverter, A.D., Dalrymple, B.P., 2009. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput Biol.* 5 (5), pp.e1000382.

Ishida, M. and Moore, G.E., 2013. The role of imprinted genes in humans. *Mol Aspects Med.* 34 (4), pp.826-40.

Jahn, T., Seipel, P., Urschel, S., Peschel, C. and Duyster, J., 2002. Role for the adaptor protein Grb10 in the activation of Akt. *Mol Cell Biol*. 22 (4), pp.979-91.

Joyce, C.A., Sharp, A., Walker, J.M., Bullman, H. and Temple, I.K., 1999. Duplication of 7p12.1-p13, including GRB10 and IGFBP1, in a mother and daughter with features of Silver-Russell syndrome. *Hum. Genet.* 105 (3), pp.273-80.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., Shoemaker, D.D., 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*. 302 (5653), pp.2141-4.

Ju, Z., Wells, M.C., Heater, S.J., Walter, R.B., 2007. Multiple tissue gene expression analyses in Japanese medaka (Oryzias latipes) exposed to hypoxia. *Comp Biochem Physiol C Toxicol Pharmacol.* 145 (1), pp.134-44.

Kabir, N.N. and Kazi, J.U., 2014. Grb10 is a dual regulator of receptor tyrosine kinase signalling. *Mol Biol Rep*. 41 (4), pp.1985-92.

Kahles, A., Ong, C.S., Zhong, Y., Rätsch, G., 2016. *SplAdder*: identification, quantification and testing of alternative splicing events from RNA-Seq data. *Bioinformatics.* 32 (12), pp.1840-7.

Kahvejian, A., Quackenbush, J., Thompson, J.F., 2008. What would you do if you could sequence everything? *Nat Biotechnol*. 26 (10), pp.1125-33.

Kan, Z., Rouchka, E.C., Gish, W.R. & States, D.J., 2001. Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.* 11 (5), pp.889-900.

Kan, Z., States, D., Gish, W., 2002. Selecting for functional alternative splices in ESTs. *Genome Res*. 12 (12), pp.1837-45.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., 2011. Spatio-temporal transcriptome of the human brain. *Nature*. 478 (7370), pp.483-9.

Kazi, J.U. and Rönnstrand, L., 2013. FLT3 signals via the adapter protein Grb10 and overexpression of Grb10 leads to aberrant cell proliferation in acute myeloid leukaemia. *Mol Oncol.* 7 (3), pp.402-18.

Kebache, S., Ash, J., Annis, M.G., Hagan, J., Huber, M., Hassard, J., Stewart, C.L., Whiteway, M., Nantel, A., 2007. Grb10 and active Raf-1 kinase promote Bad-dependent cell survival. *J Biol Chem.* 282 (30), pp.21873-83.

Kermicle, J.L., 1970. Dependence of the R-mottle aleurone phenotype in maize on mode of sexual transmission. *Genetics.* 66 (1357), pp.69-85.

Kostka, D., and Spang, R., 2004. Finding disease specific alterations in the co-expression of genes. *Bioinformatics.* 20 (Suppl 1), pp.194-9.

Kotzot, D., Schmitt, S., Bernasconi, F., Robinson, W.P., Lurie, I.W., Ilyina, H., Mehes, K., Hamel, B.C., Otten, B.J., Hergersberg, M., Werder, E, Schoenle, E., Schinzel, A., 1995. Uniparental disomy 7 in Silver-Russell syndrome and primordial growth retardation. *Hum. Mol. Genet*. 4 (4), pp.583-7.

Krawczak, M., Reiss, J., Cooper, D.N., 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and

consequences. *Hum Genet*. 90 (1-2), pp.41-54.

Langfelder, P. and Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 9 (1), pp.559.

Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature.* 409 (6822), pp.860-921.

Laviola, L., Giorgino, F., Chow, J.C., Baquero, J.A., Hansen, H., Ooi, J., Zhu, J., Riedel, H. and Smith, R.J., 1997. The adapter protein GRB10 associates preferentially with the insulin receptor as compared with the IGF-I receptor in mouse fibroblasts. *J. Clin. Invest*. 99 (5), pp.830-7.

Lee, C., Atanelov, L., Modrek, B., Xing, Y., 2003. ASAP: the Alternative Splicing Annotation Project. *Nucleic Acids Res*., 31 (1) pp.101-5.

Lee, G. and Hall, J.C., 2001. Abnormalities of male-specific FRU protein and serotonin expression in the CNS of fruitless in Dorsophila. *J Neurosci.* 21 (2), pp.513-26.

Lenth, R.V., 2001. Some Practical Guidelines for Effective Sample Size Determination. *The American Statistician*. 55 (3), 187-93.

Li, B. and Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12, pp.323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R.; 1000 Genome Project Data Processing Subgroup., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25 (16), pp.2078-9.

Li, J.J., Jiang, C.R., Brown, J.B., Huang, H., Bickel, P.J., 2011. Sparse linear modeling of next-generation mRNA sequencing (RNA-Seq) data for isoform discovery and abundance estimation. *Proc Natl Acad Sci U S A*. 108 (50), pp.19867-72.

Li, L., Li, X., Zhu, Y., Zhang, M., Yin, D., Lu, J., Liu, F., Wang, C., Jia, W., 2013.

Growth receptor binding protein 10 inhibits glucose-stimulated insulin release from pancreatic β-cells associated with suppression of the insulin/insulin-like growth factor-1 signalling pathway. *Clin Exp Pharmacol Physiol*. 40 (12), pp.841-7.

Li and Lu L., 2005. Epidermal Growth Factor-induced Proliferation Requires Down-regulation of Pax6 in Corneal Epithelial Cells. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*. 280 (13), pp.12988-95.

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 30 (7), pp.923-30.

Licatalosi, D.D. and Darnell, R.B., 2006. Splicing regulation in neurologic disease. *Neuron*. 52 (1), pp.93-101.

Lipscombe, D., 2005. Neuronal proteins custom designed by alternative splicing. Curr. Opin. *Neurobiol*. 15 (3), pp.358-63.

Lindzey, G., Winston, H., Manosevitz, M., 1961. Social dominance in inbred mouse strains. *Nature.* 191 (4787), pp.474-6.

Liu, F. and Roth, R.A., 1995. Grb-IR: a SH2-domain-containing protein that binds to the insulin receptor and inhibits its function. *Proc Natl Acad Sci*. 92 (22), pp.10287–10291.

Liu, Y., Zhou, J., and White, K.P., 2014. RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*. 30 (3), pp.301-4.

Lopes, A.M., Machado, J.A.T. and Galhano, A.M., 2017. Computational Comparison and Visualization of Viruses in the Perspective of Clinical Information. *Interdisciplinary Sciences: Computational Life Sciences*. doi:10.1007/s12539-017-0229-4.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15 (12), pp.550.

Lyon, M.F. and Glenister, P.H., 1977. Factors affecting the observed number of young

resulting from adjacent-2 disjunction in mice carrying a translocation. *Genetics Research.* 29 (1), pp.83-92.

Mano, H., Ohya, K., Miyazato, A., Yamashita, Y., Ogawa, W., Inazawa, J., Ikeda, U., Shimada, K., Hatake, K., Kasuga, M., Ozawa, K. and Kajigaya, S., 1998. Grb10/GrbIR as an in vivo substrate of Tec tyrosine kinase. *Genes Cells.* 3 (7), pp.431-41.

Marco, A., Konikoff, C., Karr, T.L., Kumar, S., 2009. Relationship between gene coexpression and sharing of transcription factor binding sites in Drosophila melanogaster. *Bioinformatics*. 25 (19), pp.2473-7.

Masuo, Y., Imai, T., Shibato, J., Hirano, M., Jones, O.A., Maguire, M.L., Satoh, K., Kikuchi, S., Rakwal, R., 2009. Omic analyses unravels global molecular changes in the brain and liver of a rat model for chronic Sake (Japanese alcoholic beverage) intake. *Electrophoresis.* 30 (8), pp.1259-75.

Matlin, A.J., Clark, F. & Smith, C.W., 2005. Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6 (5), pp.386-98.

McCann, J.A., Zheng, H., Islam, A., Goodyer, C.G. and Polychronakos, C., 2001. Evidence against GRB10 as the gene responsible for Silver-Russell syndrome. Biochem. Biophys. *Res. Commun.* 286 (5), pp.943-8.

McKee, A.E., Minet, E., Stern, C., Riahi, S., Stiles, C.D., Silver, P.A., 2005. A genome-wide in situ hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev*. Biol. 5, pp.14.

Merkin, J., Russell, C., Chen, P., Burge, C.B., 2012. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science*. 338 (6114), pp.1593-9.

Mezlini, A.M., Smith, E.J., Fiume, M., Buske, O., Savich, G.L., Shah, S., AparicioS., Chiang D. Y., Goldenberg A, Brudno, M., 2013. iReckon: simultaneous isoform discovery and abundance estimation from RNA-seq data. *Genome Res*. 23 (3), pp.519-29.

Miller, J.A., Oldham, M.C., Geschwind, D.H., 2008. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci*. 28 (6), pp.1410–20.

Modrek, B. and Lee, C., 2002. A genomic view of alternative splicing. *Nat. Genet.* 30 (1), pp.13-19.

Modrek, B. and Lee, C.J., 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genet.* 34 (2), pp.177-80.

Modrek, B., Resch, A., Grasso, C., Lee, C., 2001. Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res*. 29 (13), pp.2850-9.

Mokbel, N., Hoffman, N.J., Girgis, C.M., Small, L., Turner, N., Daly, R.J., Cooney, G. J. and Holt, L.J., 2014. Grb10 Deletion Enhances Muscle Cell Proliferation, Differentiation and GLUT4 Plasma Membrane Translocation. *Journal of Cellular Physiology.* 229 (11). pp.1753-64.

Monaco, G., van Dam, S., Casal Novo Ribeiro, J.L. Larbi, A., de Magalhães, J.P., 2015. A comparison of human and mouse gene co-expression networks reveals conservation and divergence at the tissue, pathway and disease levels. *BMC Evol Biol.* 15 (1), pp.259.

Monk, D., Arnaud, P., Frost, J., Hills, F.A., Stanier, P., Feil, R. and Moore, G.E., 2009. Reciprocal imprinting of human GRB10 in placental trophoblast and brain: evolutionary conservation of reversed allelic expression. *Human Molecular Genetics.* 18 (16), pp.3066-74.

Monk, D., Wakeling, E.L., Proud, V., Hitchins, M., Abu-Amero, S.N., Stanier, P., Preece, M.A. and Moore G.E., 2000. Duplication of 7p11.2-p13, Including GRB10, in Silver-Russell Syndrome. *Am*. *J. Hum. Genet.* 66 (1), pp.36-46.

Monzón-Sandoval, J., Castillo-Morales, A., Urrutia, A.O., & Gutierrez, H., 2016. Modular reorganization of the global network of gene regulatory interactions during

perinatal human brain development. *BMC Developmental Biology*. 16, pp.13.

Morrione, A., 2000. Grb10 proteins in insulin-like growth factor and insulin receptor signaling (review). *Int J Mol Med.* 5 (2), pp.151-4.

Morrione, A., 2003. Grb10 adapter protein as regulator of insulin-like growth factor receptor signalling. *J Cell Physiol*. 197 (3), pp.307-11.

Morrione, A., Valentinis, B., Resnicoff, M., Xu, S. and Baserga, R., 1997. The role of mGrb10alpha in insulin-like growth factor I-mediated growth. *J. Biol. Chem*. 272 (42), pp.26382-7.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 5 (7), pp.621-8.

Moutoussamy, S., Renaudie, F., Lago, F., Kelly, P.A. and Finidori, J., 1998. Grb10 identified as a potential regulator of growth hormone (GH) signalling by cloning of GH receptor target proteins. *J Biol Chem*. 273 (26), pp.15906-12.

Murdaca, J., Treins., C., Monthouël-Kartmann, M.N., Pontier-Bres, R., Kumar, S., Van Obberghen, E. and Giorgetti-Peraldi. S., 2004. Grb10 prevents *Nedd4*-mediated vascular endothelial growth factor receptor-2 degradation. *J Biol Chem*. 279 (25), pp.26754-61.

Nagy, A., Gertensenstein, M., Vintersten, K. and Behringer, R., 2003. Manipulating the Mouse Embryo: A Laboratory Manual (3rd edn). *Cold Spring Harbor Laboratory Press*. New York: *Cold Spring Harbor*. 764, pp.198-200, 268-271.

Nantel, A., Huber, M. and Thomas, D.Y., 1999. Localization of endogenous Grb10 to the mitochondria and its interaction with the mitochondrial-associated Raf-1 pool. *J Biol Chem*. 274 (50), pp.35719-24.

Nantel, A., Mohammad-Ali, K., Sherk, J., Posner, B.I. and Thomas, D.Y., 1998. Interaction of the Grb10 Adapter Protein with the Raf1 and MEK1 Kinases. *J. Biol. Chem.* 273 (17), pp.10475-84.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature.* 489 (7414), pp.83-90.

Newman, S.W., 1999. The medial extended amygdala in male reproductive behaviour. A node in the mammalian social behaviour network. *Ann N Y Acad Sci.* 877, pp.242-57.

NCBI, 2014. [Online]. Available from http://www.ncbi.nlm.nih.gov/gene/?term=Grb10 [Accessed date 20th December 2014].

NCBI, 2018. [Online]. Available from https://www.ncbi.nlm.nih.gov/gene/207 [Accessed date 20th January 2018].

Nilsson, D., Gunasekera, K., Mani, J., Osteras, M., Farinelli, L., Baerlocher, L., Roditi, I., Ochsenreiter, T., 2010. Spliced leader trapping reveals widespread alternative splicing patterns in the highly dynamic transcriptome of Trypanosoma brucei. *PLoS Pathog*. 6 (8), pp. e1001037.

Nurtdinov, R.N., Artamonova, I.I., Mironov, A.A. & Gelfand, M.S., 2003. Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.* 12 (11), pp.1313-20.

Obayashi, T. and Kinoshita, K., 2011. COXPRESdb: a database to compare gene coexpression in seven model animals. *Nucleic Acids Res*. 39 (Database issue), pp.D1016-22.

O'Connell, L.A., Hofmann, H.A., 2011. Genes, hormones, and circuits: an integrative approach to study the evolution of social behavior. *Front Neuroendocrinol.* 32 (3), pp.320-35.

Oldham, M.C., Horvath, S., Geschwind, D.H., 2006. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci U S A*. 103 (47), pp.17973-8.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., Geschwind, D.H., 2008. Functional organization of the transcriptome in human brain. *Nat Neurosci*. 11 (11), pp.1271-82.

O'Neill, T.J., Rose, D.W., Pillay, T.S., Hotta, K., Olefsky, J.M. and Gustafson, T.A., 1996. Interaction of a GRB-IR splice variant (a human GRB10 homolog) with the insulin and insulin-like growth factor I receptors. Evidence for a role in mitogenic signalling. *J Biol Chem.* 271 (37), pp.22506-13.

Ooi, J., Yajnik, V., Immanuel, D., Gordon, M., Moskow, J.J., Buchberg, A.M. and Margolis, B., 1995. The cloning of Grb10 reveals a new family of SH2 domain proteins. *Oncogene*. 10 (8), pp.1621-30.

Opitz, L., Salinas-Riester, G., Grade, M., Jung, K., Jo, P., Emons, G., Ghadimi, B.M., Beissbarth, T. and Gaedcke, J., 2010. Impact of RNA degradation on gene expression profiling. *BMC Med Genomics*. 3, pp.36.

Orban, P.C., Chui, D., Marth, J.D., 1992. Tissue- and site-specific DNA recombination in transgenic mice. *PNAS.* 89 (15), pp.6861-5.

Oshlack, A. and Wakefield, M.J., 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct.* 4, pp.14.

Ozsolak, F., Milos, P.M., 2011. RNA sequencing: advances, challenges and opportunities. *Nature reviews Genetics*. 12 (2), pp.87-98.

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 40 (12), pp.1413-5.

Pandey, A., Duan, H., Di Fiore, P.P. and Dixit, V.M., 1995. The Ret receptor protein tyrosine kinase associates with the SH2-containing adapter protein Grb10. *J Biol Chem*. 270 (37), pp.21461-3.

Patro, R., Duggal, G., Love, M.I, Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 14 (4),

pp.417-19.

Patro, R., Mount, S.M., Kingsford, C., 2014. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. *Nat Biotech*. 32 (5), pp.462-4.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., Salzberg, S.L., 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 33 (3), pp.290-5.

Pierson, E., GTEx Consortium, Koller, D., Battle, A., Mostafavi, S., Ardlie, K.G., Getz, G., Wright, F.A., Kellis, M., Volpi, S., Dermitzakis, E.T., 2015. Sharing and specificity of co-expression networks across 35 human tissues. *PLoS Comput Biol*. 11 (5), pp.e1004220.

Plasschaert, R.N. and Bartolomei, M.S., 2015. Tissue-specific regulation and function of Grb10 during growth and neuronal commitment. *Proc Natl Acad Sci U S A.* 112 (22), pp.6841-7.

Ploner Alexander, 2015. Heatplus: Heatmaps with row and/or column covariates and colored clusters. (version 2.24.0) [R package]. Available from: https://github.com/alexploner/Heatplus.

Ponomarev, I., Wang, S., Zhang, L., Harris, R.A., Mayfield, R.D., 2012. Gene coexpression networks in human brain identify epigenetic modifications in alcohol dependence. *J Neurosci*. 32 (5), pp.1884-97.

Porte, B., Chatelain, C., Hardouin, J., Derambure, C., Zerdoumi, Y., Hauchecorne, M., Dupré, N., Bekri, S., Gonzalez, B., Marret, S., Cosette, P., Leroux, P., 2017. Proteomic and transcriptomic study of brainmicrovessels in neonatal and adult mice. *PLoS One*. 12 (1), pp.e0171048.

Ramos, F.J., Langlais, P.R., Hu, D., Dong, L.Q. and Liu, F., 2006. Grb10 mediates insulin-stimulated degradation of the insulin receptor: a mechanism of negative regulation. *Am J Physiol Endocrinol Metab*. 290 (6), pp.E1262-6.

Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., and Betel, D., 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*. 14 (9), pp.R95.

Reese, M.G., Hartzell, G., Harris, N.L., Ohler, U., Abril, J.F. and Lewis, S.E., 2000. Genome annotation assessment in Drosophila melanogaster. *Genome Res*. 10 (4), pp.483-501.

Relógio, A., Ben-Dov, C., Baum, M., Ruggiu, M., Gemund, C., Benes, V., Darnell, R.B., Valcárcel, J., 2005 Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J Biol Chem.* 280 (6), pp.4779-84.

Reznik, E. & Sander, C., 2015. Extensive decoupling of metabolic genes in cancer*. PLoS Comput Biol.* 11 (5), pp.e1004176.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 43 (7), pp.e47.

Riquelme Medina, I. and Lubovac-Pilav, Z., 2016. Gene Co-Expression Network Analysis for Identifying Modules and Functionally Enriched Pathways in Type 1 Diabetes. *PLoS One*. 11 (6), pp.e0156006.

Roberts, A. and Pachter, L., 2013. Streaming fragment assignment for real-time analysis of sequencing experiments. *Nat Methods.* 10 (1), pp.71-3.

Roberts, A., Pimentel, H., Trapnell, C., Pachter, L., 2011. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. 27 (17), pp.2325-9.

Robinson, M.D. and Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 11 (3) p.R25.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 26

(1), pp.139-40.

Rozengurt Enrique, 1992. Growth factors and cell proliferation. *Current Opinion in Cell Biology*. 4 (2), pp.161-5.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. 230 (4732), pp.1350-4.

Sandelowski M., 1995. Sample size in qualitative research. *Research in Nursing & Health*. 18 (2), pp.179–83.

Sanz, L.A., Chamberlain, S., Sabourin, J.C., Henckel, A., Magnuson, T., Hugnot, J.P., Feil, R. and Arnaud, P., 2008. A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. *EMBO J*. 27 (19), pp.2523-32.

Saris, C.G., Horvath, S., van Vught, P.W., van Es, M.A., Blauw, H.M., Fuller, T.F., Langfelder, P., DeYoung, J., Wokke, J.H., Veldink, J.H., 2009. Weighted gene coexpression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics*. 10 (1), pp.405.

Sarna, J.R., Dyck, R.H., Whishaw, I.Q., 2000. The Dalila effect: C57BL6 mice barber whiskers by plucking. *Behav Brain Res.* 108 (1), pp.39-45.

Schafer, S., Miao, K., Benson, C.C., Heinig, M., Cook, S.A., Hubner, N., 2015. Alternative Splicing Signatures in RNA-seq Data: Percent Spliced in (PSI). *Curr Protoc Hum Genet*. 87, pp.11.16.1-14.

Schrader Franz, 1921. The chromosomes of Pseudococcus nipæ. *Biological Bulletin*. 40 (5), pp.259-69.

Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G.G., Owen-Hughes, T., Blaxter, M., Barton, G.J., 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. *RNA*. 22 (6), pp.839-51.

Schurch, N.J., Schofield, P., Gierli´nski, M., Cole, C., Simpson, G.G., Hughes, T.O., Blaxter, M., and Barton, G.J., 2015. Evaluation of tools for differential gene expression analysis by RNA-seq on a 48 biological replicate experiment. *Arxiv.* pp.1-31.

Shen, S., Park, J.W., Lu, Z.X., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A*. 111 (51), pp. E5593-601.

Shi, Y. and Jiang, H., 2013. rSeqDiff: detecting differential isoform expression from RNA-Seq data using hierarchical likelihood ratio test. *PLoS One*. 8 (11), pp.e79448.

Shin, C. & Manley, J.L., 2004. Cell signalling and the control of pre-mRNA splicing. *Nat. Rev. Mol. Cell Biol.* 5 (9), pp.727-38.

Singh, D., Orellana, C.F., Hu, Y., Jones, C.D., Liu, Y., Chiang, D.Y., Liu, J., Prins, J.F., 2011. FDM: a graph-based statistical method to detect differential transcription using RNA-seq data. *Bioinformatics.* 27 (19), pp.2633-40.

Smith, F.M., Garfield, A.S. and Ward, A., 2006. Regulation of growth and metabolism by imprinted genes. *Cytogenet Genome Res.* 113 (1-4), pp.279-91.

Smith, F.M., Holt, L.J., Garfield, A.S., et al. (13 co-authors)., 2007. Mice with a disruption of the imprinted Grb10 gene exhibit altered body composition, glucose homeostasis, and insulin signalling during postnatal life. *Mol Cell Biol.* 27 (16), pp.5871-86.

Soneson, C. and Delorenzi, M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*. 14, pp.91.

Srivastava, A., Sarkar, H., Gupta, N., Patro, R., 2016. RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics*. 32 (12), pp.i192-i200.

Stead, J.D., Neal, C., Meng, F., Wang, Y., Evans, S., Vazquez, D.M., Watson, S.J., Akil, H., 2006. Transcriptional profiling of the developing rat brain reveals that the most dramatic regional differentiation in gene expression occurs postpartum. *J*

Neurosci. 26 (1), pp.345-53.

Stein, E., Cerretti, D.P. and Daniel, T.O., 1996. Ligand activation of ELK receptor tyrosine kinase promotes its association with Grb10 and Grb2 in vascular endothelial cells. *J Biol Chem*. 271 (38), pp.23588-93.

Stein, E.G., Gustafson, T.A. and Hubbard, S.R., 2001. The BPS domain of Grb10 inhibits the catalytic activity of the insulin and IGF1 receptors. *FEBS Lett*. 493 (2-3), pp.106-11.

Sternberg, N., Hamilton, D., Austin, S., Yarmolinsky, M., Hoess, R., 1981. Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harb Symp Quant Biol*. 45, pp.297–309.

Sterner, K.N., Weckle, A., Chugani, H.T., Tarca, A.L., Sherwood, C.C., Hof, P.R., Kuzawa, C.W., Boddy, A.M., Abbas, A., Raaum, R.L., 2012. Dynamic gene expression in the human cerebral cortex distinguishes children from adults. *PLoS One*. 7 (5), pp. e37714.

Stuart, J.M., Segal, E., Koller, D., Kim, S.K., 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science*. 302 (5643), pp.249-55.

Southworth, L.K., Owen, A.B., Kim, S.K., 2009. Aging mice show a decreasing correlation of gene expression within genetic modules. *PLoS Genet*. 5 (12), pp.e1000776.

Sultan, M., Schulz, M.H., Richard, H., et al. (16 co-authors)., 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*. 321 (5891), pp.956-60.

Tandonnet, S., Torres, T.T., 2017. Traditional *versus* 3′ RNA-seq in a non-model species. *Genomics Data*. 11, pp.9-16.

Thanaraj, T.A., Clark, F. & Muilu, J., 2003. Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* 31 (10), pp.2544-52.

Thanaraj, T.A., Stamm, S., Clark, F., Riethoven, J.J., Le Texier, V., Muilu, J., 2004. ASD: The Alternative Splicing Database. *Nucleic Acids Res*. 32 (90001), pp.D64-69.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., Pachter, L., 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 31 (1), pp.46-53.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R, Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protocols*. 7 (3), pp.562-78.

Tesson, BM., Breitling, R., Jansen, RC., 2010. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. *BMC bioinformatics*. 11 (1), pp.497.

Tezuka, N., Brown, A.M. and Yanagawa, S., 2007. GRB10 binds to LRP6, the Wnt co-receptor and inhibits canonical Wnt signalling pathway. *Biochem Biophys Res Commun.* 356 (3), pp.648-54.

Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., Calos, M.P., 2001. Site-specific genomic integration in mammalian cells mediated by phage ΦC31 integrase. *Mol Cell Biol*. 21 (12), 3926-34.

Torkamani, A., Dean, B., Schork, N.J., Thomas, E.A., 2010. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res*. 20 (4), pp.403-12.

Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 98 (9), pp.5116-21.

Ule, J., Darnell, R.B., 2006. RNA binding proteins and the regulation of neuronal synaptic plasticity. *Curr. Opin. Neurobiol.* 16 (1), pp.102-10.

Ureche, O.N., Ureche, L., Henrion, U., Strutz-Seebohm, N., Bundis, F., Steinmeyer,

K., Lang, F., Seebohm, G., 2009. Differential modulation of cardiac potassium channels by Grb adaptor proteins. *Biochem Biophys Res Commun*. 384 (1), pp.28-31.

Urschel, S., Bassermann, F., Bai, R.Y., Münch, S., Peschel, C., Duyster, J., 2005. Phosphorylation of Grb10 regulates its interaction with 14-3-3. *J Biol Chem*. 280 (17), pp.16987-93.

Usadel, B., Obayashi, T., Mutwil, M., Giorgi, F.M., Bassel, G.W., Tanimoto, M., Chow, A., Steinhauser, D., Persson, S., Provart, N.J., 2009. Co-expression tools for plant opportunities for hypothesis generation and caveats. *Plant Cell Environ*. 32 (12), pp.1633-51.

van Dam, S., Craig, T., de Magalhães, J.P., 2015. GeneFriends: a human RNA-seq-based gene and transcript co-expression database. *Nucleic Acids Res*. 43 (Database issue), pp.D1124-32.

van Dam, S., Võsa, U., van der Graaf, A., Franke, L., de Magalhães, J.P., 2017. Gene co-expression analysis for functional classification and gene-disease predictions. *Brief Bioinform*. 19 (4), pp.575-92.

Vecchione, A., Marchese, A., Henry, P., Rotin, D. and Morrione, A., 2003. The *Grb10/Nedd4* complex regulates ligand-induced ubiquitination and stability of the insulin-like growth factor I receptor. *Mol Cell Biol*. 23 (9), pp.3363-72.

Vélez, P., Schwartz, A.B., Iyer, S.R., Warrington, A., Fadool, D.A., 2016. Ubiquitin ligase *Nedd4*-2 modulates Kv1.3 current amplitude and ion channel protein targeting. *J Neurophysiol*. 116 (2), pp.671-85.

Villella, A., Gailey, D.A., Berwald, B., Ohshima, S., Barnes, P.T., Hall, J.C., 1997. Extended reproductive roles of the fruitless gene in Drosophila melanogaster revealed by behavioral analysis of new fru mutants. *Genetics*. 147 (3), pp.1107-30.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., Geschwind, D.H., 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature*. 474 (7351), pp.380-4.

Wang, F., Zhu, J., Zhu, H., Zhang, Q., Lin, Z., Hu, H., 2011. Bidirectional control of social hierarchy by synaptic efficacy in medial prefrontal cortex. *Science.* 334 (6056), pp.693-7.

Wang, J., Dai, H., Yousaf, N., Moussaif, M., Deng, Y., Boufelliga, A., Swamy, O. R., Leone, M. E. and Riedel, H., 1999. Grb10, a Positive, Stimulatory Signaling Adapter in Platelet Derived Growth Factor BB-, Insulin-Like Growth Factor I-, and Insulin-Mediated Mitogenesis. *Mol Cell Biol*. 19 (9), pp.6217-28.

Wang, L., Balas, B., Christ-Roberts, C.Y., Kim, R.Y., Ramos, F.J., Kikani, C.K., Li, C., Deng, C., Reyna, S., Musi, N., Dong, L.Q., DeFronzo, R.A., Liu, F., 2007. Peripheral disruption of the Grb10 gene enhances insulin signaling and sensitivity in vivo. *Mol Cell Biol*. 27 (18), pp.6497-505.

Wang, W., Qin, Z., Feng, Z., Wang, X., Zhang, X., 2013. Identifying differentially spliced genes from two groups of RNA-seq samples. *Gene.* 518 (1), pp.164-70.

Watson, F.L., Püttmann-Holgado, R., Thomas, F., Lamar, D.L., Hughes, M., Kondo, M., Rebel, V.I., Schmucker, D., 2005. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science.* 309 (5742), pp.1874-78.

Weickert, C.S., Elashoff, M., Richards, A.B., Sinclair, D., Bahn, S., et al., 2009. Transcriptome analysis of male-female differences in prefrontal cortical development. *Molecular psychiatry.* 14 (6), pp.558-61.

Weyn-Vanhentenryck, S.M., Feng, H., Ustianenko, D., Duffié, R., Yan, Q., Jacko, M., Martinez, J.C., Goodwin, M., Zhang, X., Hengst, U., Lomvardas, S., Swanson, M.S., Zhang, C., 2018. Precise temporal regulation of alternative splicing during neural development. *Nat Commun.* 9 (1), pp.2189.

Whitley, S.K., Horne, W.T., Kolls, J.K., 2016. Research Techniques Made Simple: Methodology and Clinical Applications of RNA Sequencing. *J Invest Dermatol.* 136 (8), pp.e77-82.

Wick, K.R., Werner, E.D., Langlais, P., Ramos, F. J., Dong, L.Q., Shoelson, S.E.

and Liu, F., 2003. Grb10 inhibits insulin-stimulated insulin receptor substrate (IRS)-phosphatidylinositol 3-kinase/Akt signalling pathway by disrupting the association of IRS-1/IRS-2 with the insulin receptor. *J Biol Chem*. 278 (10), pp.8460-7.

Wikimedia Commons contributors, 'File:Protein GRB10 PDB 1nrv.png', *Wikimedia Commons, the free media repository,* 20 March 2015, UTC. Available from https://commons.wikimedia.org/w/index.php?title=File:Protein_GRB10_PDB_1nrv.png&oldid=154048952 [accessed 14th June 2018].

Wood, S.H., Craig, T., Li, Y., Merry, B., de Magalhães, J.P., 2013. Whole transcriptome sequencing of the aging rat brain reveals dynamic RNA changes in the dark matter of the genome. *Age (Dordr)*. 35 (3), pp.763-76.

Xiao, R., Tang, P., Yang, B., Huang, J., Zhou, Y., Shao, C., Li, H., Sun, H., Zhang, Y., Fu, X.D., 2012. Nuclear matrix factor hnRNP U/SAF-A exerts a global control of alternative splicing by regulating U2 snRNP maturation. *Mol Cell.* 45 (5), pp.656-68.

Xu, X., Zhan, M., Duan, W., et al. (15 co-authors)., 2007. Gene expression atlas of the mouse central nervous system: impact and interactions of age, energy intake and gender. *Genome Biol.* 8 (11), pp.R234.

Xu, Q., Modrek, B., Lee, C., 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res*. 30 (17), pp.3754-66.

Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y., Jiang, L., Cai, Z., Sun, H., Zhang, K., Zhang, Y., Chen, J., Fu, X.D., 2013. Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell*. 152 (1-2), pp.82-96.

Xue, Y., Zhou, Y., Wu, T., Zhu, T., Ji, X., Kwon, Y.S., Zhang, C., Yeo, G., Black, D.L., Sun, H., Fu, X.D., Zhang, Y., 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell*. 36 (6), pp.996-1006.

Yang, S., Deng, H., Zhang, Q., Xie, J., Zeng, H., Jin, X., Ling, Z., Shan, Q., Liu,

M., Ma, Y., Tang, J., Wei, Q., 2016. Amelioration of Diabetic Mouse Nephropathy by Catalpol Correlates with Down-Regulation of Grb10 Expression and Activation of Insulin-Like GrowthFactor 1 / Insulin-Like Growth Factor 1 Receptor Signaling. *PLoS One*. 11 (3), pp.e0151857.

Yamasaki-Ishizaki, Y., Kayashima, T., Mapendano, C.K., Soejima, H., Ohta, T., Masuzaki, H., Kinoshita, A., Urano, T., Yoshiura, K., Matsumoto, N., Ishimaru, T., Mukai, T., Niikawa, N., Kishino, T., 2007. Role of DNA methylation and histone H3 lysine 27 methylation in tissue-specific imprinting of mouse Grb10. *Mol Cell Biol*. 27 (2), pp.732-42.

Yeo, G., Holste, D., Kreiman, G., Burge, C.B., 2004. Variation in alternative splicing across human tissues. *Genome Biol*. 5 (10), pp.R74.

Yu, H., Luscombe, N.M., Qian, J., Gerstein, M., 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet*. 19 (8), pp.422-7.

Yoshihashi, H., Maeyama, K., Kosaki, R., Ogata, T., Tsukahara, M., Goto, Y., Hata, J., Matsuo, N., Smith, R.J. and Kosaki, K., 2000. Imprinting of human GRB10 and its mutations in two patients with Russell-Silver syndrome. *Am. J. Hum. Genet.* 67 (2), pp.476-82.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., et al. (14 co-authors) et al., 2015. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*. 347 (6226), pp.1138-42.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., 2013. Integrated systems approach identifies genetic nodes and networks in late-onset al.zheimer's disease. *Cell*. 153 (3), pp.707-20.

Zhang, J., Lu, K., Xiang, Y., Islam, M., Kotian, S., Kais, Z., Lee, C., Arora, M., Liu, H.W., Parvin, J.D., 2012A. Weighted frequent gene co-expression network mining to identify genes involved in genome stability. *PLoS Comput Biol*. 8 (8), pp.1002656.

Zhang, J., Zhao, J., Jiang, W., Shan, X., Yang, X., Gao, J., 2012B. Conditional gene manipulation: Cre-ating a new biological era* . J *Zhejiang Univ Sci B.* 13 (7), pp.511-24.

Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F., Black, D.L., Kharchenko, P.V., Sharp, P.A., Walsh, C.A., 2016. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell*. 166 (5), pp.1147-1162.

Zhang, Y., Chen, K., Sloan, S.A., et al. (17 co-authors)., 2014. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34 (36), pp.11929-47.

Zhang, Z. and Wang, W., 2014. RNA-Skim: a rapid method for RNA-Seq quantification at transcript level. *Bioinformatics*. 30 (12), pp.283-92.

Zhao, S., Fung-Leung, WP., Bittnerm A, Ngo, K., Liu, X., 2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS*. 9 (1), pp.e78644.

Zheng, C.L., Kwon, Y.S., Li, H.R., Zhang, K., Coutinho-Mansfield, G., Yand, C., Nair, T.M., Gribskov, M., Fu, X.D., 2005. MAASE: an alternative splicing database designed for supporting splicing microarray applications. *RNA*. 11 (12), pp.1767-76.

Zhou J., Zhao S., Dunker A.K., 2018. Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signaling and Regulation. *J Mol Biol.* 430 (16), pp.2342-2359.

Supplementary material

Appendix S2.1.

DISSECTION RECORD (E18.5)

Date of dissection: / /201

Plug date: / /201

Time of sacrifice of the pregnant mouse (min)

Pup No.	Subcortical areas R/part L/part RNAlater	Cortex RNAlater	Collecting time of brain tissue in RNAlater /minute	Tail clip Genotype	Sex	Total time between killing the pregnant mouse and collecting pups' brain sample	Given No. For Male only	Genotype	Notes
			:			min			
			:			min			
			:			min			
			:			min			
			:			min			

Appendix S2.2.

DISSECTION RECORD (1wk old) / (1 month old) / (3 months old) / (6 months old)

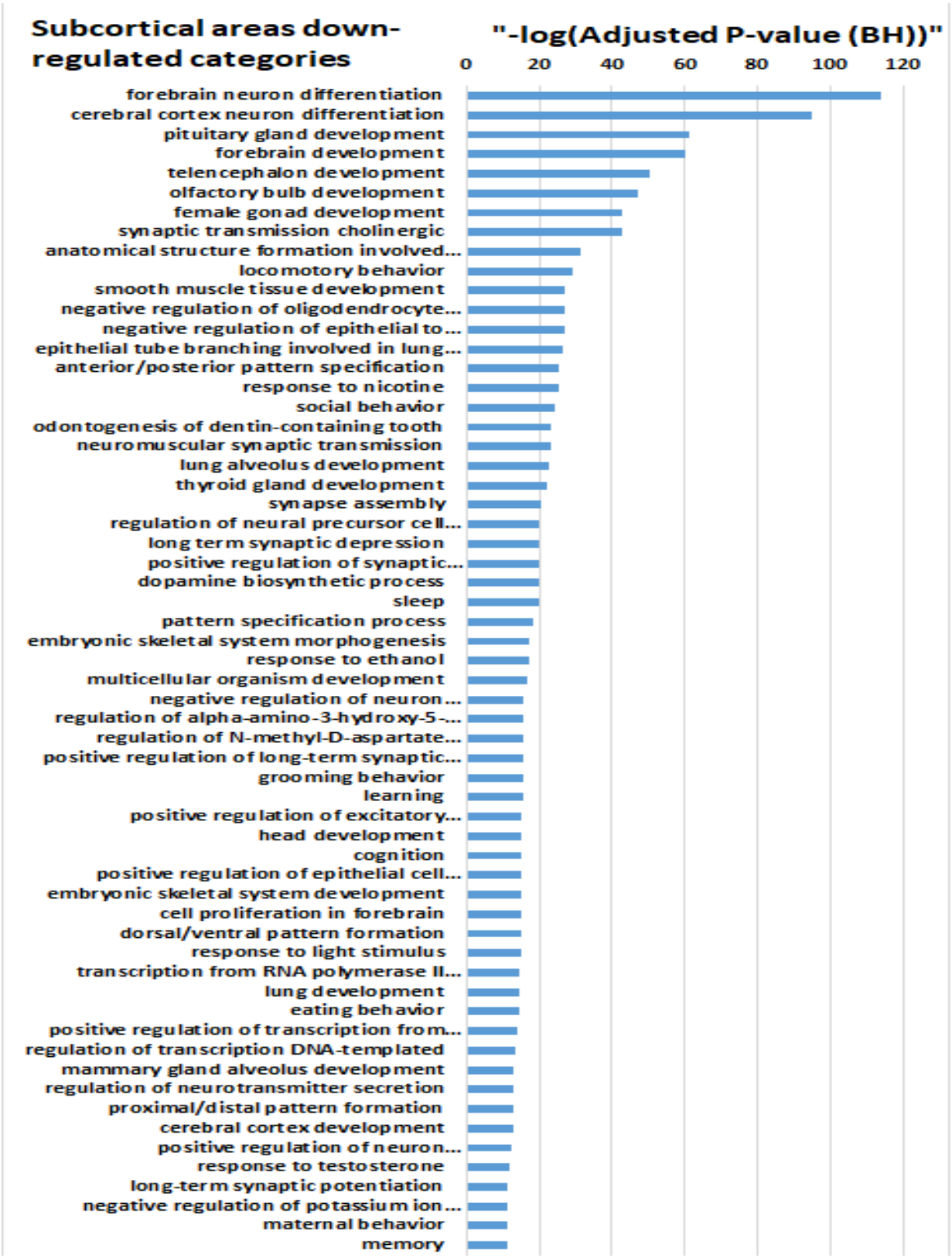
Date of dissection: / /201 Date of birth: / /201

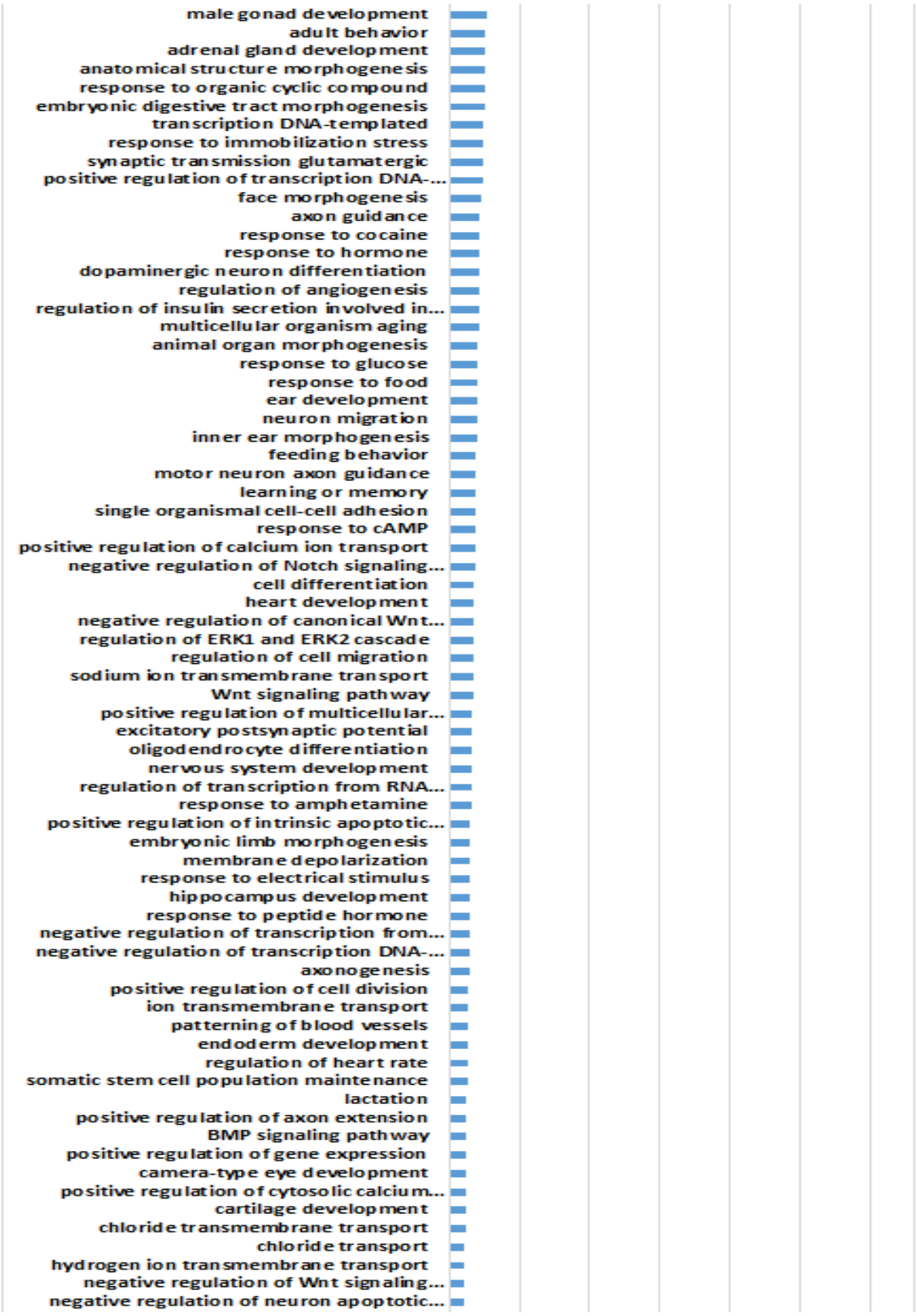
No	Killing time of the mouse /minute	Subcortical areas R/part L/part RNAlater	Cortex RNAlater	Time of brain sample collection in RNA/min	Tail clip Genotype	Sex	Total time between killing the mouse & collecting the brain sample	Given No.	Genotype	Notes
	:			:			min			
	:			:			min			
	:			:			min			
	:			:			min			

Table S2.3. Information of dissected mice. Details of mice used for this project such as their type (WT, mutant), date of birth, age, parental identity number and their date of birth. * F1 (first generation): the exact number and their date of birth were not provided.

No.	Mouse type	Birth Date	Age	Paternal I.D.&BirthDate	Maternal I.D.&BirthDate
1	*Grb10KO*$^{+/p}$	02/04/2015	18.5	KO 3 (19/12/2014)	WT 4 (19/12/2014)
2	*Grb10KO*$^{+/p}$	20/05/2015	18.5	KO 5 (09/01/2015)	WT 4 (02/02/2015)
3	*Grb10KO*$^{+/p}$	20/05/2015	18.5	KO 5 (09/01/2015)	WT 4 (02/02/2015)
4	WT	02/04/2015	18.5	KO 5 (18/12/2014)	WT 3 (19/12/2014)
5	WT	02/04/2015	18.5	KO 5 (18/12/2014)	WT 3 (19/12/2014)
6	WT	20/05/2015	18.5	KO 5 (09/01/2015)	WT 4 (02/02/2015)
7	*Grb10KO*$^{+/p}$	03/05/2015	1W	KO 4 (19/12/2014)	WT 1 (19/12/2014)
8	*Grb10KO*$^{+/p}$	06/05/2015	1W	KO 6 (09/01/2015)	WT 3 (08/01/2015)
9	*Grb10KO*$^{+/p}$	06/05/2015	1W	KO 6 (09/01/2015)	WT 3 (08/01/2015)
10	WT	06/05/2015	1W	KO 4 (19/12/2014)	WT 2 (08/01/2015)
11	WT	06/05/2015	1W	KO 4 (19/12/2014)	WT 2 (08/01/2015)
12	WT	20/06/2015	1W	KO 5 (09/01/2015)	WT 2 (03/04/2015)
13	*Grb10KO*$^{+/p}$	09/05/2015	1M	KO 4 (09/01/2015)	WT 2 (08/01/2015)
14	*Grb10KO*$^{+/p}$	09/05/2015	1M	KO 4 (09/01/2015)	WT 2 (08/01/2015)
15	*Grb10KO*$^{+/p}$	09/05/2015	1M	KO 4 (09/01/2015)	WT 2 (08/01/2015)
16	WT	16/03/2015	1M	KO 5 (27/09/2014)	WT F1
17	WT	16/03/2015	1M	KO 5 (27/09/2014)	WT F1
18	WT	09/05/2015	1M	KO 4 (09/01/2015)	WT 2 (08/01/2015)
19	*Grb10KO*$^{+/p}$	02/10/2014	3M	KO 4 (11/05/2014)	WT F1
20	*Grb10KO*$^{+/p}$	02/10/2014	3M	KO 4 (11/05/2014)	WT F1
21	*Grb10KO*$^{+/p}$	02/11/2014	3M	KO 3 (05/06/2014)	WT F1
22	WT	02/10/2014	3M	KO 4 (11/05/2014)	WT F1
23	WT	11/10/2014	3M	KO 3 (05/06/2014)	WT F1
24	WT	02/11/2014	3M	KO 3 (05/06/2014)	WT F1
25	*Grb10KO*$^{+/p}$	16/08/2014	6M	KO 1 (04/04/2014)	WT 3 (17/03/2014)
26	*Grb10KO*$^{+/p}$	23/11/2014	6M	KO 2 (05/06/2014)	WT F1
27	*Grb10KO*$^{+/p}$	23/11/2014	6M	KO 2 (05/06/2014)	WT F1
28	WT	16/08/2014	6M	KO 1 (04/04/2014)	WT 3 (17/03/2014)
29	WT	27/10/2014	6M	KO 4 (11/05/2014)	WT F1
30	WT	23/11/2014	6M	KO 2 (05/06/2014)	WT F1

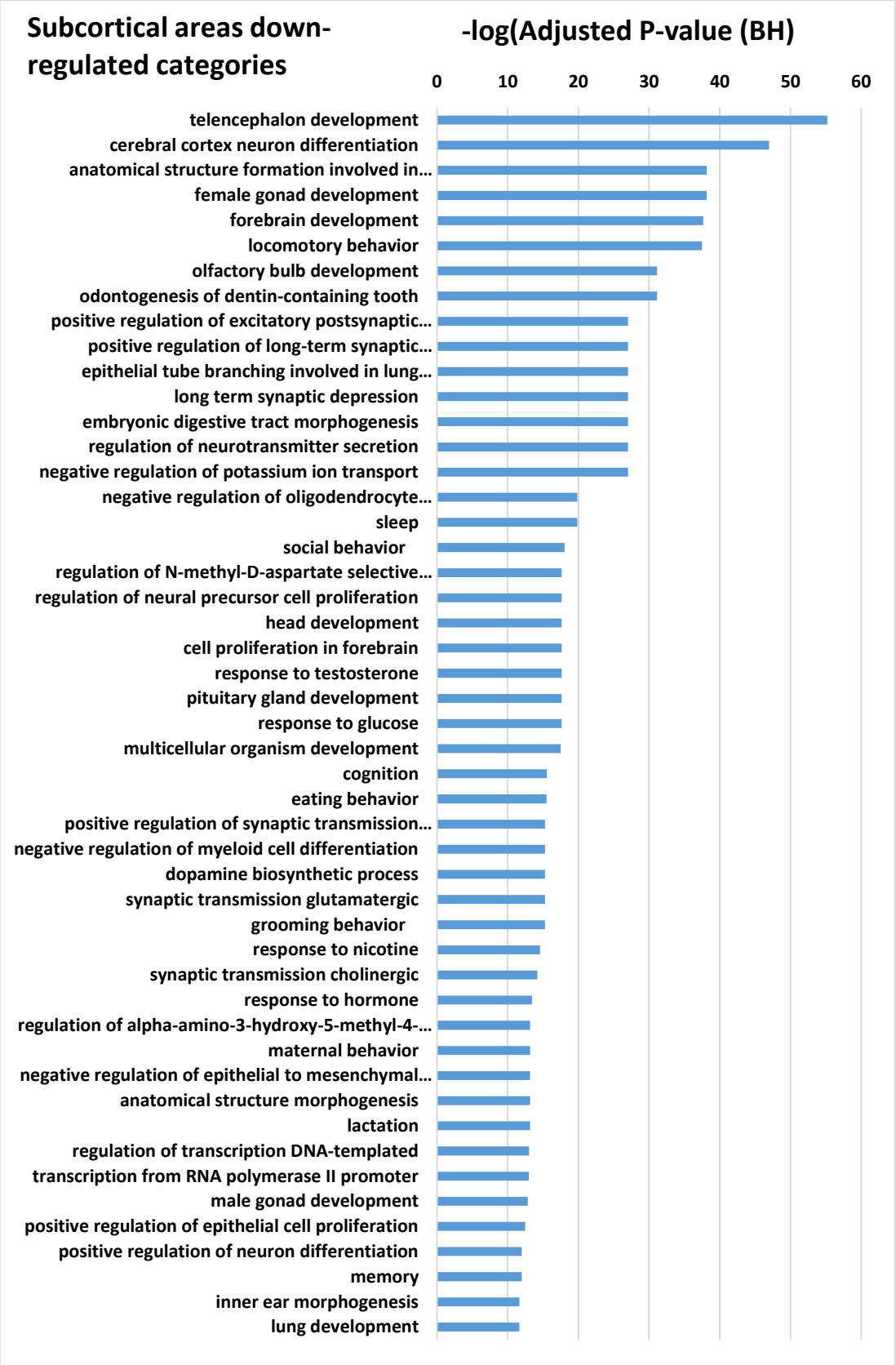
Figure S3.1. Bar plot showing the number of significantly enriched GO terms. Gene ontology enrichment analysis of biological processes for down-regulated differentially expressed genes in the subcortical areas.




male gonad development
adult behavior
adrenal gland development
anatomical structure morphogenesis
response to organic cyclic compound
embryonic digestive tract morphogenesis
transcription DNA-templated
response to immobilization stress
synaptic transmission glutamatergic
positive regulation of transcription DNA-...
face morphogenesis
axon guidance
response to cocaine
response to hormone
dopaminergic neuron differentiation
regulation of angiogenesis
regulation of insulin secretion involved in...
multicellular organism aging
animal organ morphogenesis
response to glucose
response to food
ear development
neuron migration
inner ear morphogenesis
feeding behavior
motor neuron axon guidance
learning or memory
single organismal cell-cell adhesion
response to cAMP
positive regulation of calcium ion transport
negative regulation of Notch signaling...
cell differentiation
heart development
negative regulation of canonical Wnt...
regulation of ERK1 and ERK2 cascade
regulation of cell migration
sodium ion transmembrane transport
Wnt signaling pathway
positive regulation of multicellular...
excitatory postsynaptic potential
oligodendrocyte differentiation
nervous system development
regulation of transcription from RNA...
response to amphetamine
positive regulation of intrinsic apoptotic...
embryonic limb morphogenesis
membrane depolarization
response to electrical stimulus
hippocampus development
response to peptide hormone
negative regulation of transcription from...
negative regulation of transcription DNA-...
axonogenesis
positive regulation of cell division
ion transmembrane transport
patterning of blood vessels
endoderm development
regulation of heart rate
somatic stem cell population maintenance
lactation
positive regulation of axon extension
BMP signaling pathway
positive regulation of gene expression
camera-type eye development
positive regulation of cytosolic calcium...
cartilage development
chloride transmembrane transport
chloride transport
hydrogen ion transmembrane transport
negative regulation of Wnt signaling...
negative regulation of neuron apoptotic...


negative regulation of cell proliferation
cellular response to glucose stimulus
cellular response to calcium ion
negative regulation of myeloid cell...
response to estradiol
response to steroid hormone
cardiac muscle contraction
negative regulation of BMP signaling...
embryonic cranial skeleton morphogenesis
glycogen metabolic process
response to ischemia
neurotransmitter transport
protein localization to plasma membrane
regulation of membrane potential
transport
cell development
glucose metabolic process
activation of protein kinase activity
somitogenesis
blood vessel development
neuron projection development
transmembrane transport
epithelial cell differentiation
skeletal muscle tissue development
ion transport
positive regulation of protein kinase...
positive regulation of...
brain development
cell proliferation
response to activity
skeletal muscle cell differentiation
cellular response to drug
response to calcium ion
sodium ion transport
cell-matrix adhesion
positive regulation of peptidyl-tyrosine...
activation of GTPase activity
negative regulation of sequence-specific...
cellular response to retinoic acid
positive regulation of protein binding
angiogenesis
Notch signaling pathway
negative regulation of cell migration
positive regulation of protein...
positive regulation of JNK cascade
protein oligomerization
response to drug
MAPK cascade
regulation of ion transmembrane transport
cell chemotaxis
activation of cysteine-type endopeptidase...
female pregnancy
response to mechanical stimulus
cellular response to interleukin-1
positive regulation of cell growth
chemokine-mediated signaling pathway
response to glucocorticoid
regulation of cell proliferation
cation transmembrane transport
cellular response to hypoxia
regulation of cell shape
osteoblast differentiation
post-embryonic development
calcium ion transport
peptidyl-tyrosine phosphorylation
multicellular organism growth
neuropeptide signaling pathway
positive regulation of ERK1 and ERK2...
positive regulation of protein kinase B...
response to hypoxia
palate development

Figure S3.2. Bar plot showing the number of significantly enriched GO terms. GO categories of biological process for down-regulated differentially expressed genes in the subcortical areas at 3 months old mice.




multicellular organism aging
neuromuscular synaptic transmission
response to organic cyclic compound
response to cocaine
negative regulation of neuron...
pattern specification process
transcription DNA-templated
learning
cellular response to glucose stimulus
regulation of insulin secretion involved in...
face morphogenesis
adult behavior
feeding behavior
axon guidance
positive regulation of transcription from...
long-term synaptic potentiation
regulation of transcription from RNA...
cerebral cortex development
embryonic limb morphogenesis
regulation of angiogenesis
negative regulation of canonical Wnt...
positive regulation of calcium ion transport
regulation of heart rate
response to cAMP
animal organ morphogenesis
cell differentiation
regulation of cell migration
neuron migration
excitatory postsynaptic potential
negative regulation of Notch signaling...
camera-type eye development
response to immobilization stress
proximal/distal pattern formation
synapse assembly
hippocampus development
response to electrical stimulus
positive regulation of transcription DNA-...
embryonic cranial skeleton morphogenesis
response to activity
endoderm development
glycogen metabolic process
cardiac muscle contraction
somitogenesis
dopaminergic neuron differentiation
response to steroid hormone
BMP signaling pathway
positive regulation of cytosolic calcium...
learning or memory
Wnt signaling pathway
skeletal muscle cell differentiation
response to ischemia
response to ethanol
negative regulation of transcription from...
cell development
nervous system development
negative regulation of cell proliferation
positive regulation of...
somatic stem cell population maintenance
ion transmembrane transport
oligodendrocyte differentiation
embryonic skeletal system development
response to amphetamine
brain development


negative regulation of neuron apoptotic...
heart development
negative regulation of BMP signaling...
positive regulation of cell growth
response to peptide hormone
sodium ion transmembrane transport
patterning of blood vessels
positive regulation of protein binding
skeletal muscle tissue development
positive regulation of intrinsic apoptotic...
negative regulation of transcription DNA-...
response to calcium ion
protein oligomerization
negative regulation of Wnt signaling...
chemokine-mediated signaling pathway
chloride transport
regulation of ion transmembrane transport
dorsal/ventral pattern formation
chloride transmembrane transport
regulation of cell proliferation
epithelial cell differentiation
response to mechanical stimulus
positive regulation of protein kinase...
cell chemotaxis
negative regulation of sequence-specific...
palate development
activation of protein kinase activity
calcium ion transport
in utero embryonic development
cell proliferation
axonogenesis
cellular response to calcium ion
activation of GTPase activity
cellular response to interleukin-1
activation of cysteine-type endopeptidase...
single organismal cell-cell adhesion
positive regulation of GTPase activity
ion transport
MAPK cascade
positive regulation of JNK cascade
regulation of membrane potential
transport
cellular response to drug
negative regulation of cell migration
neuron differentiation
post-embryonic development
regulation of gene expression
response to estradiol
anterior/posterior pattern specification
Notch signaling pathway
cellular response to hypoxia
neuropeptide signaling pathway
positive regulation of ERK1 and ERK2...
response to hypoxia
cellular response to tumor necrosis factor
potassium ion transmembrane transport
sodium ion transport
regulation of apoptotic process
calcium ion transmembrane transport
peptidyl-tyrosine phosphorylation
potassium ion transport
neuron projection development
small GTPase mediated signal transduction
aging

Figure S3.3. Multidimensional scaling plot of the normalised data samples of WT mice. Red colour represents tissue samples of the subcortical areas while blue colour represents tissue samples of the cortex. Different developmental age group of WT mice represented by certain symbols: E18.5 ■ ■, 1W ▲ ▲, 1M ● ●, 3M ◆ ◆ and 6M ▼ ▼ .

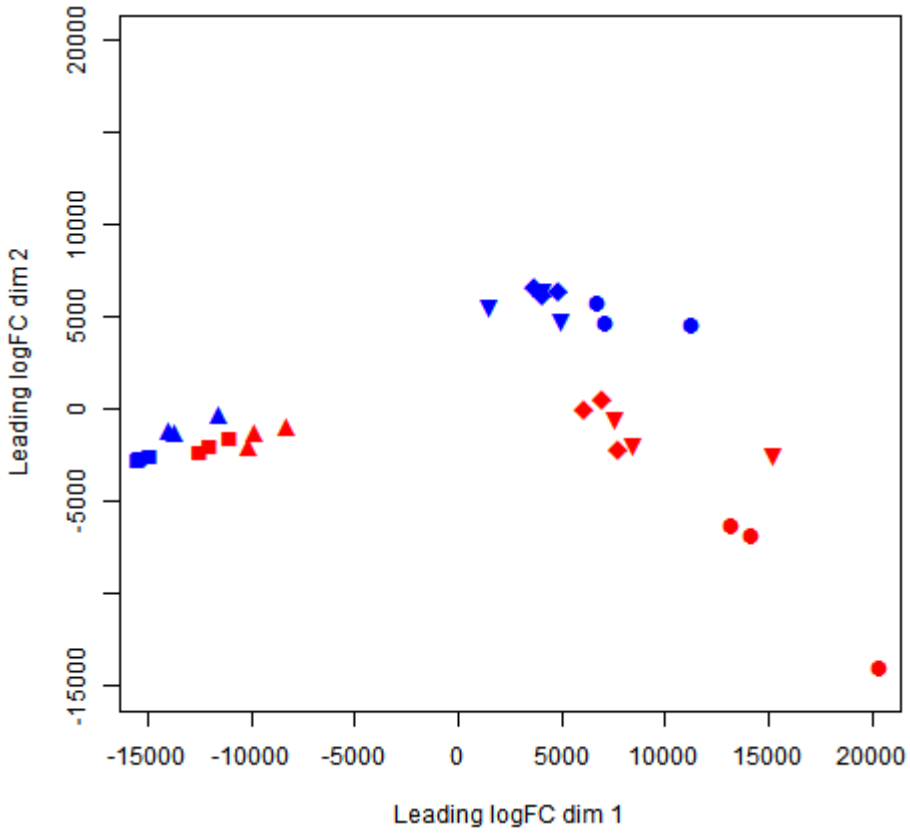


Table S3.1. Shared GO terms of biological process (A1, A2) and cellular component (B) between the subcortical areas and the cortex for both up- and down-regulated differentially expressed genes. *Shared GO terms are in bold colour.

A1. Biological process

Subcortical a.-Down regu		Subcortical a.-Upregulated	Cortex-Downregul	Cortex-Upregulated	All GO terms _ Subcortical areas		All GO terms_Cortex
GO:0001666	GO:0072659	GO:0001666	0	GO:0008285	**GO:0008285**	GO:0001568	**GO:0008285**
GO:0007218	GO:0042391	GO:0007218		GO:0010628	**GO:0010628**	GO:0001756	**GO:0010628**
GO:0007631	GO:0006810	GO:0007631		GO:0014070	**GO:0014070**	GO:0031175	**GO:0014070**
GO:0042755	GO:0006006	GO:0042755		GO:0042493	**GO:0042493**	GO:0055085	**GO:0042493**
GO:0045471	GO:0032147	GO:0045471		GO:0045471	**GO:0045471**	GO:0007519	**GO:0045471**
GO:0021879	GO:0001568	GO:0055093		GO:0045893	**GO:0045893**	GO:0030855	**GO:0045893**
GO:0021895	GO:0001756	GO:0001731		GO:0048468	**GO:0048468**	GO:0006811	**GO:0048468**
GO:0021983	GO:0031175	GO:0015986		GO:0010862	GO:0001666	GO:0014068	GO:0010862
GO:0030900	GO:0055085	GO:0006754		GO:0045786	GO:0007218	GO:0045860	GO:0045786
GO:0021537	GO:0007519	GO:0015992		GO:0033138	GO:0007631	GO:0007420	GO:0033138
GO:0021772	GO:0030855			GO:0043408	GO:0042755	GO:0008283	GO:0043408
GO:0014070	GO:0006811			GO:0060395	GO:0021879	GO:0014823	GO:0060395
GO:0045893	GO:0014068			GO:0007623	GO:0021895	GO:0035914	GO:0007623
GO:0010628	GO:0045860			GO:0045907	GO:0021983	GO:0035690	GO:0045907
GO:0008285	GO:0007420			GO:0042981	GO:0030900	GO:0051592	GO:0042981
GO:0048468	GO:0008283			GO:0043066	GO:0021537	GO:0006814	GO:0043066
GO:0042493	GO:0014823			GO:0043065	GO:0021772	GO:0007160	GO:0043065
GO:0007271	GO:0035914			GO:0006412	GO:0007271	GO:0050731	GO:0006412
GO:0008585	GO:0035690			GO:0008284	GO:0008585	GO:0090630	GO:0008284
GO:0048646	GO:0051592				GO:0048646	GO:0043433	
GO:0007626	GO:0006814				GO:0007626	GO:0071300	
GO:0010719	GO:0007160				GO:0010719	GO:0001525	
GO:0048715	GO:0050731				GO:0048715	GO:0032092	
GO:0048745	GO:0090630				GO:0048745	GO:0007219	
GO:0060441	GO:0043433				GO:0060441	GO:0030336	
GO:0009952	GO:0071300				GO:0009952	GO:0001934	
GO:0035094	GO:0001525				GO:0035094	GO:0046330	
GO:0035176	GO:0032092				GO:0035176	GO:0051259	
GO:0042475	GO:0007219				GO:0042475	GO:0000165	
GO:0007274	GO:0030336				GO:0007274	GO:0034765	
GO:0048286	GO:0001934				GO:0048286	GO:0060326	
GO:0030878	GO:0046330				GO:0030878	GO:0006919	
GO:0007416	GO:0051259				GO:0007416	GO:0007565	
GO:0030431	GO:0000165				GO:0030431	GO:0009612	
GO:0042416	GO:0034765				GO:0042416	GO:0071347	
GO:0051968	GO:0060326				GO:0051968	GO:0030307	
GO:0060292	GO:0006919				GO:0060292	GO:0070098	
GO:2000177	GO:0007565				GO:2000177	GO:0051384	
GO:0007389	GO:0009612				GO:0007389	GO:0042127	
GO:0048704	GO:0071347				GO:0048704	GO:0098655	
GO:0007275	GO:0030307				GO:0007275	GO:0008360	
GO:0045665	GO:0070098				GO:0045665	GO:0071456	
GO:0007625	GO:0051384				GO:0007625	GO:0001649	
GO:1900273	GO:0042127				GO:1900273	GO:0009791	
GO:2000310	GO:0098655				GO:2000310	GO:0006816	
GO:2000311	GO:0008360				GO:2000311	GO:0018108	
GO:0007612	GO:0071456				GO:0007612	GO:0035264	
GO:0009416	GO:0001649				GO:0009416	GO:0070374	
GO:0009953	GO:0009791				GO:0009953	GO:0051897	
GO:0021846	GO:0006816				GO:0021846	GO:0060021	
GO:0048706	GO:0018108				GO:0048706	GO:0055093	
GO:0050679	GO:0035264				GO:0050679	GO:0001731	
GO:0050890	GO:0070374				GO:0050890	GO:0015986	
GO:0060322	GO:0051897				GO:0060322	GO:0006754	
GO:2000463	GO:0060021				GO:2000463	GO:0015992	

A2. Biological process

Subcortical a.-Down regu	Subcortical a.-Upregulated	Cortex-Downregul	Cortex-Upregulated	All GO terms _ Subcortical areas	All GO terms_Cortex
GO:0006366	GO:0045746			GO:0006366 GO:0001975	
GO:0030324	GO:0051928			GO:0030324 GO:2001244	
GO:0045944	GO:0030154			GO:0045944 GO:0030326	
GO:0006355	GO:0007507			GO:0006355 GO:0051602	
GO:0009954	GO:0090090			GO:0009954 GO:0051899	
GO:0046928	GO:0070372			GO:0046928 GO:0021766	
GO:0060749	GO:0030334			GO:0060749 GO:0043434	
GO:0021987	GO:0035725			GO:0021987 GO:0000122	
GO:0045666	GO:0016055			GO:0045666 GO:0045892	
GO:0033574	GO:0040018			GO:0033574 GO:0007409	
GO:0060291	GO:0007399			GO:0060291 GO:0051781	
GO:0042711	GO:0048709			GO:0042711 GO:0034220	
GO:0043267	GO:0060079			GO:0043267 GO:0001569	
GO:0007613	GO:0006357			GO:0007613 GO:0007492	
GO:0008584	GO:0001975			GO:0008584 GO:0002027	
GO:0009653	GO:2001244			GO:0009653 GO:0035019	
GO:0030325	GO:0030326			GO:0030325 GO:0007595	
GO:0030534	GO:0051602			GO:0030534 GO:0045773	
GO:0048557	GO:0051899			GO:0048557 GO:0030509	
GO:0006351	GO:0021766			GO:0006351 GO:0043010	
GO:0035249	GO:0043434			GO:0035249 GO:0007204	
GO:0035902	GO:0000122			GO:0035902 GO:0051216	
GO:0060325	GO:0045892			GO:0060325 GO:1902476	
GO:0007411	GO:0007409			GO:0007411 GO:0006821	
GO:0009725	GO:0051781			GO:0009725 GO:1902600	
GO:0042220	GO:0034220			GO:0042220 GO:0030178	
GO:0045765	GO:0001569			GO:0045765 GO:0043524	
GO:0071542	GO:0007492			GO:0071542 GO:0071333	
GO:0010259	GO:0002027			GO:0010259 GO:0071277	
GO:0061178	GO:0035019			GO:0061178 GO:0045638	
GO:0009887	GO:0007595			GO:0009887 GO:0032355	
GO:0009749	GO:0045773			GO:0009749 GO:0048545	
GO:0032094	GO:0030509			GO:0032094 GO:0060048	
GO:0043583	GO:0043010			GO:0043583 GO:0030514	
GO:0001764	GO:0007204			GO:0001764 GO:0048701	
GO:0042472	GO:0051216			GO:0042472 GO:0005977	
GO:0007611	GO:1902476			GO:0007611 GO:0002931	
GO:0008045	GO:0006821			GO:0008045 GO:0006836	
GO:0016337	GO:1902600			GO:0016337 GO:0072659	
GO:0051591	GO:0030178			GO:0051591 GO:0042391	
GO:0060048	GO:0043524			GO:0045746 GO:0006810	
GO:0030514	GO:0071333			GO:0051928 GO:0006006	
GO:0048701	GO:0071277			GO:0030154 GO:0032147	
GO:0005977	GO:0045638			GO:0007507 GO:0035725	
GO:0002931	GO:0032355			GO:0090090 GO:0016055	
GO:0006836	GO:0048545			GO:0070372 GO:0040018	
				GO:0030334 GO:0007399	
				GO:0060079 GO:0048709	
				GO:0006357	

B. Cellular component

Subcortical a.-Down reg.	Subcortical a.-Upregulated	Cortex-Downregulated	Cortex-Upregulated	All GO terms_Subcortical areas	All GO terms_Cortex
GO:0030141	GO:0030141	0	GO:0005576	**GO:0005576**	**GO:0005576**
GO:0043204	GO:0043204		GO:0030141	**GO:0030141**	**GO:0030141**
GO:0016600	GO:0045263		GO:0098794	**GO:0098794**	**GO:0098794**
GO:0045202	GO:0005753		GO:0016023	GO:0043204	GO:0016023
GO:0030054	GO:0031966		GO:0005840	GO:0016600	GO:0005840
GO:0055038	GO:0043025			GO:0045202	
GO:1902711	GO:0005913			GO:0030054	
GO:0045211	GO:0005911			GO:0055038	
GO:0030425	GO:0005739			GO:1902711	
GO:0098794				GO:0045211	
GO:0014069				GO:0030425	
GO:0005892				GO:0014069	
GO:0030027				GO:0005892	
GO:0005634				GO:0030027	
GO:0090575				GO:0005634	
GO:0009898				GO:0090575	
GO:0043198				GO:0009898	
GO:0032281				GO:0043198	
GO:0046658				GO:0032281	
GO:0016235				GO:0046658	
GO:0043005				GO:0016235	
GO:0030864				GO:0043005	
GO:0042734				GO:0030864	
GO:0030424				GO:0042734	
GO:0005667				GO:0030424	
GO:0043197				GO:0005667	
GO:0030018				GO:0043197	
GO:0005791				GO:0030018	
GO:0005576				GO:0005791	
GO:0005886				GO:0005886	
GO:0043195				GO:0043195	
GO:0034707				GO:0034707	
GO:0016323				GO:0016323	
				GO:0045263	
				GO:0005753	
				GO:0031966	
				GO:0043025	
				GO:0005913	
				GO:0005911	
				GO:0005739	

Figure S6.1. GO terms for down-regulated differentially expressed genes in the liver (=73).

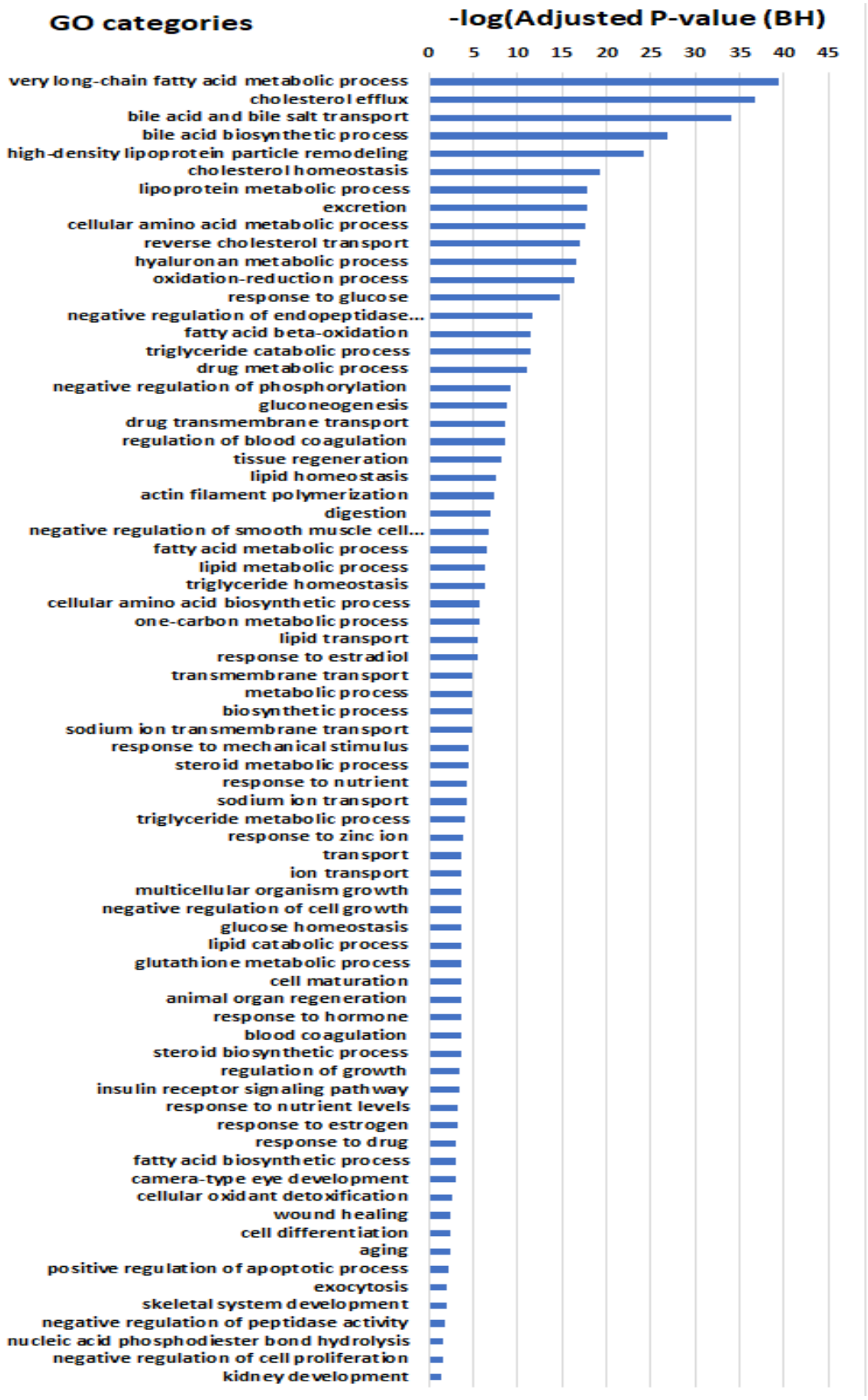


Figure S6.2. GO terms for up-regulated differentially expressed genes in the liver (=102).